Applications of Tags in Multimodal Analysis of Motion Ergonomics for Healthcare Environments

Konrad Kluwak[*]**, Marek Kulbacki**[**, ****]**, Anna Kołcz**[***]

*Wroclaw University of Science and Technology, Faculty of Electronics, Poland, konrad.kluwak@pwr.edu.pl
**Polish-Japanese Academy of Information Technology, R&D Center, Koszykowa 86, 02-008 Warsaw, Poland, mk@pja.edu.pl
***Wroclaw Medical University, Faculty of Health Science, Department of Physiotherapy, Ergonomics and Biomedical Monitoring Laboratory, Grunwaldzka 2, 50-355 Wroclaw,Poland, anna.kolcz-trzesicka@umed.wroc.pl
****DIVE IN AI, Kamienna 15/54, 53-307 Wroclaw, Poland, info@diveinai.com

Abstract: This work introduces the Extracted Tags - EXTags, a short form of data extracted from a massive amount of multimodal human motion data for efficient human motion analysis. EXTags describe the only crucial space-time features from motion data in a certain period. We demonstrate how such brief representation might be a handful in an analysis of the patient transport situation from the point of view of the ergonomics of transporting people.

Keywords: Tag Detection; Motion Analysis; Ergonomics in Healthcare

1 Introduction

The transport of patients is a routine, frequently repeated activity performed by health professionals. Statistics show that routinely introduces terrible habits, which often accelerate employee occupational diseases (OHS) associated with the nonergonomic patient transfer. In 2017 American National Institute for Occupational Safety and Health reported over 1.25 trillion dollars loss due to the OHS problems [33]. In the daily activities of professional medical staff, few tasks do not require physical assistance. Many nurses must perform physically demanding patient care tasks that expose them to an increased risk of musculoskeletal disorders (MSD) and low-back disorders (LBD). A good survey [5] covering 132 papers on the frequency of injuries and musculoskeletal disorders pain concludes that nursing aides suffer from OHS.In [24], authors report a survey of 2405 nurses with range lifting tasks and confirm that low back pain is highly prevalent among nurses and is associated with a high-level sickness absence. Alamgir et al. in [1] examined injury claims that resulted in compensation or medical costs studying 2784 injury incidents and indicating the majority of injuries in nursing homes and acute care. Terrence et al. [28] examined the relationship between the regularity of patient lifting and the occurrence of back injuries and proved that regularity is a significant causative factor in the production of low back injuries in nursing personnel. Warming et al. [32] evaluated the inter-method reliability of a registration sheet for patient handling tasks and demonstrated that the logbook was reliable for both transfer and care tasks. In his systematic review [23] of papers between 2004 and 2016, Richardson concluded that additional research is necessary to identify interventions that may reduce the high rates of injury among nurses. Skotte in [25] investigated a low-back loading during everyday patient-handling tasks, emphasizing the compression and shear forces on L4/L5 joint and measured muscle activity. Hignett in [8] proved that nursing staff had lower levels of associated postural risk in positive safety cultures. Schnibye et al. [24] observed changes in mechanical load on the low-back and assessed a significant reduction was in spinal loading.

Literature studies show that for determining correct assessment methods, correct and incorrect lifting experiments should be carried out and registered in a controlled environment. Human Motion Laboratories for functional analysis equipped to measure kinematic, kinetic parameters, and electromyography of selected muscle groups involved in lifting are suitable for such trials. Efficient processing of particular information from heterogeneous massive multimodal human motion data (structured and unstructured) requires the availability of methods to select, process, and distinguish correct and incorrect samples of motion from the data stream in a fast way. Two popular mechanisms include indexing and tagging. However, an effective tagging of multiperson and multimodal human motion data requires an abstract description. The chosen motion modality or a few selected modalities or dependencies between them could be easily compared with other data samples in motion in the same time slot and modality.

There are several ideas in the literature on how to tag multimodal data. Rasiwasia et al. [22] demonstrated the benefits of joint text and image modeling by comparing the performance of a state-of-the-art image retrieval system to his image retrieval system, using the proposed joint model. A similar method was proposed by [27]. According to [9], current image segmentation approaches are divided into four categories: merging-based, splitting-based, statistical model-based, and shot boundary classification-based. Authors claim that practical video structure analysis aims at segmenting a video and using semantic content. Atrey et al. [2] present known fusion strategies for combining multiple modalities appropriate to various tasks. In [3], the authors prove that the most appropriate is to use a taxonomy of multimodal machine learning with characteristic representation, translation, data fusion alignment, and co-learning in the multimodal world. Our proposition of EXTag representation is similar to [2] and [3] directions. We propose using several selected modalities simultaneously to represent particular motion patterns by single EXTag - Motion Tags (MT) entities. The ordered vector of Motion Tags represents multiperson motion as an instance of EXTag. We use hierarchical clustering, classification, and value derivative dynamic time warping [16] to compare normalized time series data efficiently.

To our best knowledge, current approaches consider only the analysis of artificial models or real situations without an accurate measurement methodology. In our case, the measurement concerns two people lifting supervised by professional physicians. In addition, researchers use dummy and artificial patients and simulations instead of accurate measurements of real people during activities.

Most of the existing approaches are based on a single domain, considering only the analysis of one movement feature to measure the correctness of lifting. Our main innovation introduces compound models that aggregate simultaneous activities of simple features like kinematic, kinetic, and EMG of selected body parts during motion. This experiment uses a person's motion which contains mutual relations lifter and being lifted person. The novelty starts from measurement protocol defined, which considers a lifting specialist and patient. We capture the details of the whole movement, taking into account the patient and the doctor, taking into account the kinetics, and on this basis, we build features, and these features are hierarchical. The second novelty starts from the measurement protocol defined, which considers lifting (nurse) and that the patient had been lifted.

2 Measurement Methodology

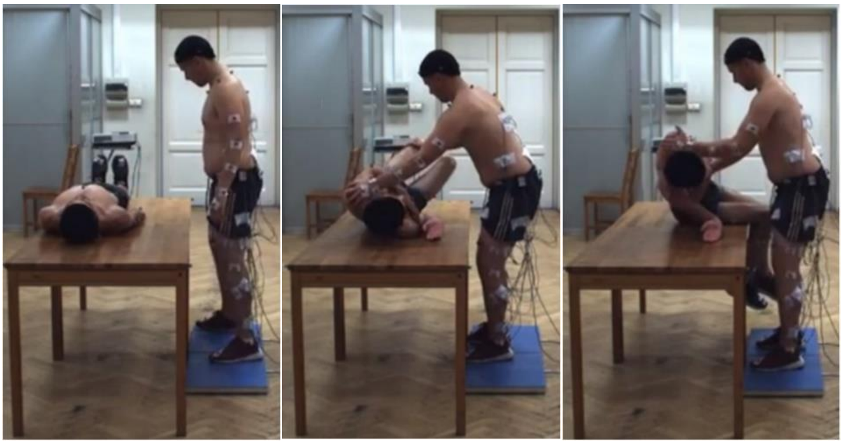
For demonstration purposes, the safe handling of the patient during the raising of a patient from lying to sitting in bed is being used. It requires two people, so it gives us complex data. The techniques of correct handling have been derived from the literature and verified by the experience of the employees of Wroclaw Medical University in Poland. Our study uses recorded multimodal information from 4 cameras, surface electromyography (sEMG), ground reaction forces (GRF), and full-body motion capture (MC). In multimodal observation, we focus on the configuration of hips, back, and knees and accompanying selected surface muscle tensions.

3 Activity Scenarios

The activity scenario is composed of three phases: 1) preparing the patient for movement, 2) turning the patient to lie down on his side, in a safe position, 3) moving the patient from a safe position to a seated position at the edge of the bed.

Correct lifting must follow the rules of the loads lifting mechanics:

- keep the spine in a neutral position (in a slight squat),
- bend of knees and hips (knees-patellas should not exceed the line of fingers),
- hold the patient close to the body at waist height (arms not extended),
- move as smoothly as possible.



(a) 1.1 (b) 2.3 (c) 3.2

Figure 1
Key phases of patient lifting (numbering as described in activity scenario procedure)

The recommended patient transfer techniques, including a patient who is heavier than the caregiver, without overloading the spine and exposing him/her to injuries, are described in detail:

1. **Prepare the patient for movement**
   1.1. Position yourself facing the bed on which the patient is lying. Take a relaxed position.
   1.2. Bend your knees slightly, position your feet freely, tighten your torso is stabilizing muscles, Put your spine in a safe position, breathe freely.

1.3. Get closer to the patient’s head, now change the position of the feet. Both feet are pointing towards the patient’s head. Forward, leg closer to the patient’s head, forward-facing, bent in the knee, bodyweight transferred to the front leg, rear leg slightly bent, leg and trunk muscles tense, spine in a safe position, free breathing.

1.4. Move the patient’s hand closer to the edge of the bed, bend it at the elbow, and then bend it at the elbow. Slide your hand under your head (pillow). Point the patient’s head towards the hand.

1.5. Move the patient’s other hand over the chest to the edge of the bed and put your hand on the bed.

1.6. Move closer to the patient’s feet, now change the position of the feet, both feet forward-facing, leg closer to the patient’s legs forward-facing, curved in the knee, bodyweight transferred to the front leg, rear leg slightly bent, leg muscles and your torso is tense, your spine is in a safe position, breath freely.

1.7. Slide your hand closer to the patient’s lying foot and bend your leg by slightly sliding it over the mattress. Then, bend the patient’s other leg in the knee.

2. **Turn the patient to lie down on his side in a safe position.**

2.1. Move closer to the patient’s torso.

2.2. Stand in front of the patient, slightly bend your knees, position your feet freely, one leg remains in front, bent in the knee, the other remains in the back, the bodyweight remains on the front leg, tighten the muscles stabilizing the torso, put the spine in a safe position, breathe freely.

2.3. Place one hand on the patient’s shoulder, do not bend the hand in the wrist, place the other hand near the patient’s hip, slowly move the body weight from the front leg to the rear leg, turning the patient to lie on his side, in a safe position.

3. **Move the patient from a safe position to a seated position at the edge of the bed.**

3.1. Change the position of your feet. Position yourself facing the bed on which the patient is lying, take a relaxed position, slightly bend your knees, put your feet forward freely, fingers slightly to the sides, tighten the torso stabilizing muscles, put your spine in a safe position, breathe freely. Move the patient’s legs towards the edge of the bed, slowly lowering them out of bed. Secure the patient’s legs with your leg.

3.2. Ask the patient to start pushing away from the bed with hand, and at the same time lightly support the patient's movement with hands so that the patient changes position from lying on the side to sitting on the bed with legs down.Support the patient’s hands on the bed, on both sides of the patient’s body. If possible, the patient’s feet should be supported (e.g., on the floor).

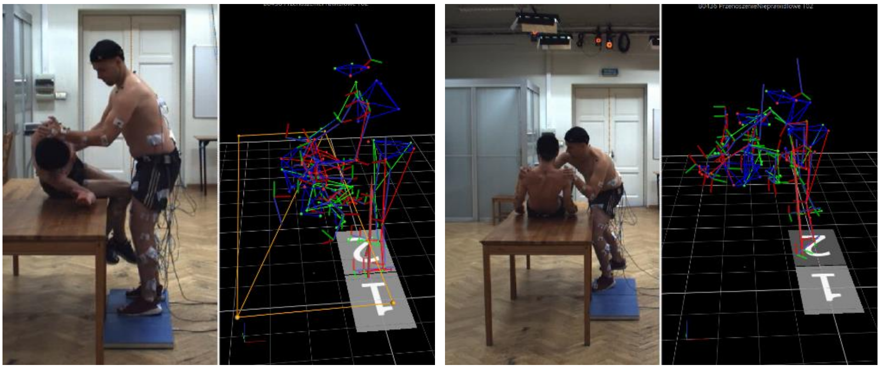
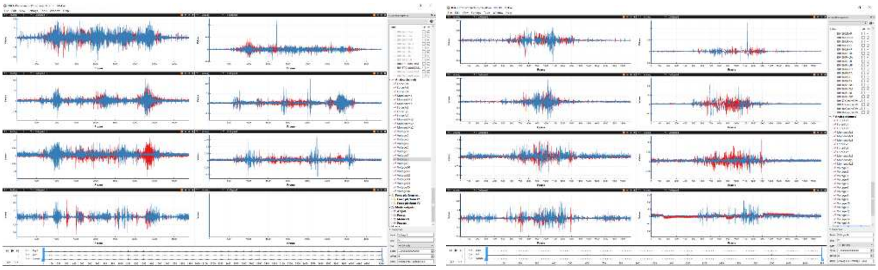


Figure 2

Multimodal Visualization of correct (left side) and incorrect (right side) patient lift with video, 3D motion, and position on the force plates



(a) Correct

(b) Incorrect

Figure 3

Examples of correct and incorrect registrations of selected muscle tensions during lifting procedure

In addition, for each actor, ten squats were registered with and without a load to perform them correctly and incorrectly. The correct motion applies to the rules written above. The incorrect does not apply. This exercise was performed as a reference to EMG measurements of patient lifting.

4 Measurement Configuration

Reference database has been registered using the system for multimodal motion acquisition in Human Motion Laboratory (HML) of the Polish-Japanese Academy of Information Technology (PJAIT) in Bytom, Poland. Laboratory provides a comprehensive environment for multimodal data acquisition, management, and analysis. It allows motion data acquisition by synchronous, simultaneous measurement and recording of video streams, muscle potentials, motion kinematics, and kinetics. It makes it possible to apply a spatio-temporal correlation between the acquired data. MX-Giganet Lab is responsible for hardware synchronization during data acquisition from base systems. Our particular multimodal recordings contain:

- **3D body motions of 2 actors** (patient and nurse) were registered based on Vicon's Motion Kinematics Acquisition and Analysis System equipped with 28 Motion Capture Camera (10 MX T40, 10 Bonita 10, and 8 Vantage 5). The system registers 39 reflective markers on each of the actors (tracked at 100 fps) based on sets for the Plug-in Gait full-body model and provides information on the location of all skeleton joints. Markers were placed on significant body segments: 4 on Head, five on Torso, 14 on the left and right side of upper limbs, and 16 on left and right side of the lower body according to Vicon specification [31].
- **Muscle potentials for nurses** by Noraxon's Dynamic Electromyography (EMG) System allows for 16-channel measurement of non-gel electrodes in compliance with the SENIAM guidelines. For registration EMG, has been defined measurement configuration (Fig. 4). EMG electrodes measure the work of muscles taking part in stabilizing the figure of the person who is lifting the patient.
- **Ground Reaction Forces (GRF)** were recorded (1000 Hz) for the nurse's body by Kistler Force Plates (Internal Amplifier), two dynamometric platforms with equal accuracy on the entire surface of the platform at 1000 Hz frequency. The system has been used to check on which leg is the more significant body pressure.
- **Video system**: using multi-camera video system for simultaneous recording from 3 Full HD (25 Hz) cameras and lossless video recording equipped with Basler's pilot piA1900-32gm/gc GigE Vision. The cameras recorded the view from the front, back, and side positions of the scene. Recordings contain information about the position of actors over time.

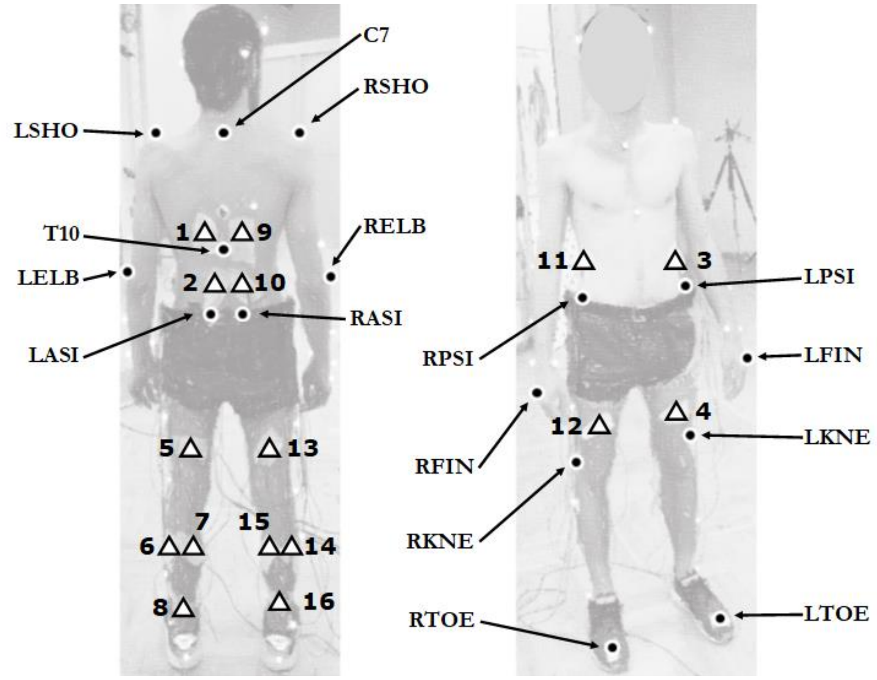


Figure 4
Lifting actor with EMG electrodes and skeleton markers configurations

Hardware synchronization and hardware calibration were described in the paper [17]. Each multimodal recording has a single C3D file containing kinematic and kinetic motion parameters, three video streams, and configuration files. In addition to patient transfer, other movements were recorded to determine maximum muscle tension in a similar situation. Two actors have been recorded performing the movements representing correct and incorrect lifting (Fig. 2). This multimodal measurement configuration has 414 distinctive motion parameters in data records, resulting in a large amount of data.

The dataset with patient lifting (DPLP) has been made public for scientific research purposes according to the initiative of Living Labs for Human Motion Analysis and Synthesis in Shared Economy Model. It is available in the resources of the R&D Center PJAIT: https://res.pja.edu.pl (accessed on 1 July 2021).

5 Extracted Tag Concept

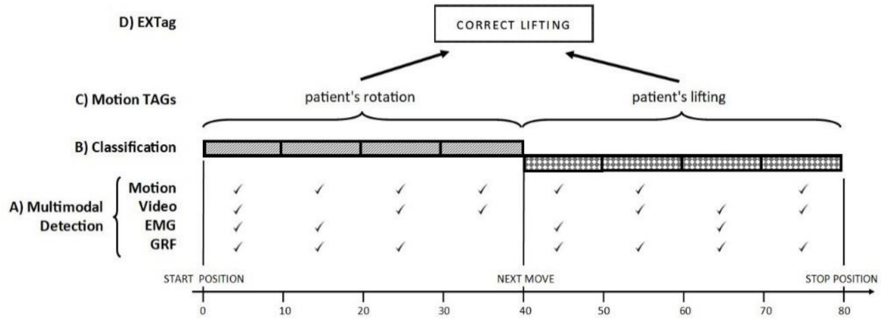


Figure 5

EXTag concept with four level of operations

To reduce data amount and faster operations, we introduce the Extracted Tag (EXTag) concept. It will allow on the performance of multimodal assessment and analysis by: (A) event detection in multimodal data, (B) ranking motion recognition, (C) Motion Tag classification, and (D) EXTag marking. These operations allow quickly performing content-based motion retrieval from complex multidimensional and multimodal space. EXTag is a part of a hierarchical representation of multimodal motion information composed of an ordered sequence of motions - Motion Tags. An example of the EXTag representation in graphical form demonstrating the patient lifting task is shown in Figure 5. EXTag is assigned to the appropriate class after the schema elements of the defined Motion Tag sequence are correctly recognized. Motion Tag is a concise, minimal representation of a given segment of motion measurement, which consists of characteristic data blocks of the selected motion. Motion Tag reference is considered as a part of multimodal measurements. In multimodal measurements, we set the characteristics of given modalities by labels: correct, incorrect. These characteristics are not the whole body configuration but the only configuration of the distinguished fragments. In multimodal motion feature space, selected features that are valid for specific Motion Tags are considered. They are applied to the Motion Tag subspace and determine whether a given Motion Tag is complete.

In addition, an order in which the given movements occur is maintained. The correct occurrence of Motion Tag determines the resulting EXTag. In this way, building a minimal representation consisting of labels and selected sets of values for specific, selected quantities during movement. These quantities are described in a more abstract way compared to the raw data recorded in the laboratory. An example of such an abstraction in our case of patient lift is the position of the arms relative to the hips, described as the mutual position of several joints relative to each other. It is not an anatomical norm but an accepted scheme that unambiguously defines the correct lifting movement in the patient lifting scheme. EXTag is the minimum representation that eliminates the need to search the entire movement space using a data fragment as a movement pattern. It is a high-level representation that does not penetrate into the detailed structure of each of the measured objects. In the case of two people's movements, the specific properties of these movements are measured in a given time. Patterns of these movements are also considered elements of measurements. The minimal representation described using EXTag consists of Motion Tags representing single motions. Every single movement is represented by specific modality values measured in time.

Tag Modelling Algorithm:

1) Define the pattern of movement - Motion Tag using motion features selected by a domain expert. Such a pattern describes the average duration of a defined movement, the properties of this movement described by changes in the kinematics and scopes of angles of joints (degrees of freedom) in the skeletal model, and at the same time changes in kinetics and muscles activities in the muscle model.

2) Select the dominant feature of the defined Motion Tag.

3) Unify a Tag for a specific modality (kinematics, EMG).

4) Collect all the characteristics of a given Motion Tag.

6 Extracted Tag Application

Motion Tag is correctly recognized only when the previously defined motion blocks represented by different modalities occur in a specific order. For that, we propose a logistic regression classifier with compressed sparse row representation that gives optimal performance. The block size depends on the specific motion description and required features selected from the corresponding measurements. Depending on the motion recording system used, the motion block may include motion capture, EMG, GRF, video, and other required modalities. The correct recognition takes place for the optimal feature vector of ordered motion states appear on multimodal data. In our multimodal data for the definition of minimal distinguishable motion representation at first, two data modalities: EMG and motion capture have been considered, which requires 70 parameters in total and results in 6 times fewer data and feature space.

In our example, we measure the motion of two people, but in general, we have P people in multiperson motion. EMG features for a person $\mathrm{P_n}, \mathrm{n} \in \{1..\mathrm{N}\}\mathrm{P_n}, \mathrm{n} \in \{1..\mathrm{N}\}$ are given by Integral Absolute Value calculated for a particular EMG_i channel represented by x_j samples in window K:

$$\boldsymbol{EMG_{P_n} = \{EMG_1, \dots, EMG_{16}\}}, \qquad \boldsymbol{EMG_i = \frac{1}{K} \sum_{j=(i-1)*N+1}^{i*K} |\, s_j \,|} \tag{1}$$

In the discussed example, for the correct lifting movement, we expect EMG modality of all muscles at a constant level and low values compared to the intense muscle load when lifting a significant weight. In the abnormal lifting movement, EMG reaches high values when the nurse overturns and lifts the patient. This stroke is identified as a threat to the spine. In order to define the optimal EMG vector, we additionally check if the tension on the back is less than on the abdomen and legs. High tension on the abdomen is identified as correct, high tension on the muscles of the squares is identified as incorrect. We do not know specific EMG values, but we rely on signal variation or a standardized pattern for a specific person based on correct and incorrect examples.

Kinematic features for person n - $\mathrm{KF}_{\mathrm{P_n}}$n - $\mathrm{KF}_{\mathrm{P_n}}$, are defined by set of cc geometric relations between specified skeleton configurations creating spatial relations in the window KK:

$$\boldsymbol{KF_{P_n} = \{KF_1, \dots, KF_c\}, \qquad KF_{\{i\}} = \left\{\sum_{j=1}^{kk}\left(w_j * v_j\right)\right\}} \tag{2}$$

where kkkk is the feature dimension of particular geometric relation, $\mathrm{v_j}\mathrm{v_j}$ particular feature and $\mathrm{w_j}\mathrm{w_j}$ normalized weight for the jj-th geometric feature.

Here we refer directly to the description of the movement from the previous chapter. In our observation, we are focused on the particular configuration of hips, back and knees. Due to the limited data, we study the rotation of the spine during lifting (hip line relative to the shoulder line), lifting work with hands close to the body. The bending of the nurse's legs in the knees when lifting reduces the lever, reducing required parameters to 32 dimension feature vector. For window KK for normalization and comparison of coherent kinematic and EMG features we use VDDTW [16] following agglomerative clustering for partitioning of mixed multimodal motion feature set into distinguished groups of primitive multimodal motions according to the distribution of generic model $\mathrm{GM_i}, , \mathrm{i} \in \{\mathrm{T}, \mathrm{F}\}\mathrm{GM_i}, , \mathrm{i} \in \{\mathrm{T}, \mathrm{F}\}$:

$$\boldsymbol{g_{i,l}\left(s_l(m)\right) = \frac{1}{\sqrt{v_l 2\pi}}\, e^{\frac{-\left[s_{l(m)} - av_l\right]^2}{2v_l}}} \tag{3}$$

For classification purposes we use similarity measure and classify multimodal motion features according to similarity measure with maximal value:

$$\boldsymbol{h(s(m)) = \underset{i \in \{1,2,\dots,K\}}{argmax}\left[P(G_i)\prod_{l=1}^{M} w_l g_{i,l}\left(s_l(m)\right)\right]} \tag{4}$$

For example, we assumed a limited set of different motions, and each group $\mathrm{G_i}\mathrm{G_i}$ has the same likelihood $\mathrm{P(G_i)} = \frac{1}{\mathrm{K}}\mathrm{P(G_i)} = \frac{1}{\mathrm{K}}$.

Conclusions

In this paper, we introduced an important problem of lifting procedure as an example of massive multiperson, multimodal motion data and elements defining the EXTags allowing for the unequivocal determination of the correctness of movement performed by medical personnel during the lifting of the patient. There was proposed reduced motion description by EXTags introduction, understood as a configuration of feature subset from a set of multimodal quantities relating to the description of motion on the high level extracted from the recorded human motion. The experiment demonstrated a significant difference between the normal and abnormal lifting characterized by limited and selected features among different motion parameters in different time slots. Such motion patterns defined for lifting a patient performed in a certain period were described by EXTag and given by specifications of multimodal motion configuration. In our study, the activity scenario and experiment description were proposed. Additionally, a multimodal measurement configuration and its impact on data quantity were presented, including the EXTag concept and its actual usability. Due to the limit of pages, the article has been significantly reduced and another paper focused on the experiment results is planned. The work will be further developed towards the analysis of human motion data for medical applications. The EXTag concept will be tested for other domains with multimodal motion data. In the future, we are considering conducting extensive experiments with different methods and models, which proved to be successful in modeling systems in various fields as presented in [4] [7] [12] [20] [34].

Acknowledgements

This study was supported by the Scientific Research from Polish-Japanese Academy of Information Technology, R&D Center and DIVE IN AI and statutory funds of Wroclaw University of Science and Technology, Faculty of Electronics and Wroclaw Medical University, Department of Physiotherapy.

References

[1] H. Alamgir, Y. Cvitkovich, S. Yu, and A. Yassi, “Work-related injury among direct care occupations in British Columbia, Canada,” Occup. Environ. Med.

[2] P. K. Atrey, M. A. Hossain, A. El Saddik, and M. S. Kankanhalli, “Multimodal fusion for multimedia analysis: A survey,” Multimed. Syst., Vol. 16

[3] T. Baltrusaitis, C. Ahuja, and L. P. Morency, “Multimodal Machine Learning: A Survey and Taxonomy,” IEEE Trans. Pattern Anal. Mach. Intell., Vol.

[4] W. Y. Cheng and C. F. Juang, “A fuzzy model with online incremental SVM and margin-selective gradient descent learning for classification problems,” IEEE Trans. Fuzzy Syst., Vol. 22, No. 2, pp. 324-337, 2014

[5] K. G. Davis and S. E. Kotowski, "Prevalence of Musculoskeletal Disorders for Nurses in Hospitals, Long-Term Care Facilities, and Home Health Care: A Comprehensive Review," Human Factors, Vol. 57, No. 5. SAGE Publications Inc., pp. 754-792, 28-Aug-2015

[6] M. Fray and S. Hignett, "Using patient handling equipment to manage mobility in and around a bed.," Loughborough University, Jan. 2015

[7] E. L. Hedrea, R. E. Precup, and C. A. Bojan-Dragos, "Results on tensor product-based model transformation of magnetic levitation systems," Acta Polytech. Hungarica, Vol. 16, No. 9, pp. 93-111, 2019

[8] S. Hignett and E. Crumpton, "Competency-based training for patient handling," Appl. Ergon., Vol. 38, No. 1, pp. 7-17, Jan. 2007

[9] W. Hu, N. Xie, L. Li, X. Zeng, and S. Maybank, "A survey on visual content-based video indexing and retrieval," IEEE Trans. Syst. Man Cybern. Part C Appl. Rev., Vol. 41, No. 6, pp. 797-819, 2011

[10] M. Jäger et al., "Lumbar-load analysis of manual patient-handling activities for biomechanical overload prevention among healthcare workers," Ann. Occup. Hyg., Vol. 57, No. 4, pp. 528-544, 2013

[11] C. Jordan, A. Luttmann, A. Theilmeier, S. Kuhn, N. Wortmann, and M. Jäger, "Characteristic values of the lumbar load of manual patient handling for the

[12] C. F. Juang, Y. Y. Lin, and R. B. Huang, "Dynamic system modeling using a recurrent interval-valued fuzzy neural network and its hardware implementation," Fuzzy Sets Syst., Vol. 179, No. 1, pp. 83-99, 2011

[13] K. Kindblom, Movement awareness and communication in patient transfer : An educational intervention. 2009

[14] K. Kjellberg, Work technique in lifting and patient transfer tasks. 2003

[15] P. Konrad, The ABC of EMG: a practical introduction to kinesiological electromyography. Noraxon USA, Inc, 2006

[16] M. Kulbacki and A. Bak, "Unsupervised Learning Motion Models Using Dynamic Time Warping," in Intelligent Information Systems 2002, Springer, 2002

[17] M. Kulbacki, J. Segen, and J. P. Nowacki, "4GAIT: Synchronized MoCap, video, GRF and EMG datasets: Acquisition, management and applications," 2014

[18] R. J. Parkinson, M. Bezaire, and J. P. Callaghan, "A comparison of low back kinetic estimates obtained through posture matching, rigid link modeling and

[19] M. Piorek, Analysis of Chaotic Behavior in Non-linear Dynamical Systems, Vol. 160, 2019

[20] C. Pozna and R. E. Precup, “Applications of signatures to expert systems modelling,” Acta Polytech. Hungarica, Vol. 11, No. 2, pp. 21-39, 2014

[21] R. E. Precup, T. A. Teban, A. Albu, A. B. Borlea, I. A. Zamfirache, and E. M. Petriu, “Evolving Fuzzy Models for Prosthetic Hand Myoelectric-Based Control,” IEEE Trans. Instrum. Meas., Vol. 69, No. 7, pp. 4625-4636, 2020

[22] N. Rasiwasia et al., “A new approach to cross-modal multimedia retrieval,” in MM’10 - Proceedings of the ACM Multimedia 2010 International Con

[23] A. Richardson, B. McNoe, S. Derrett, and H. Harcombe, “Interventions to prevent and reduce the impact of musculoskeletal injuries among nurses: A system

[24] B. Schibye, A. F. Hansen, C. T. Hye-Knudsen, M. Essendrop, M. Böcher, and J. Skotte, “Biomechanical analysis of the effect of changing patient-handling

[25] J. H. Skotte, M. Essendrop, A. F. Hansen, and B. Schibye, “A dynamic 3D biomechanical evaluation of the load on the low back during different patient-ha

[26] J. Smedley, P. Egger, C. Cooper, and D. Coggon, “Manual handling activities and risk of low back pain in nurses,” Occup. Environ. Med., Vol. 52

[27] C. G. M. Snoek and M. Worring, “Multimodal video indexing: A review of the state-of-the-art,” Multimedia Tools and Applications, Vol. 25, No. 1

[28] T. J. Stobbe, R. W. Plummer, R. C. Jensen, and M. D. Attfield, “Incidence of low back injuries among nursing personnel as a function of patient lifting

[29] M. Strohmaier, C. Körner, and R. Kern, “Understanding why users tag: A survey of tagging motivation literature and results from an empirical study,”

[30] N. Ulutasdemir and F. Tanir, “Occupational Risks of Health Professionals,” in Occupational Health, InTech, 2017

[31] Vicon Motion Systems Limited, “Full Body Modeling with Plug-in Gait,” Plug-in Gait Reference Guide, 2017

[32] S. Warming, D. H. Precht, P. Suadicani, and N. E. Ebbehøj, “Musculoskeletal complaints among nurses related to patient handling tasks and psychosocial f

[33] N. Wiggermann, “Biomechanical Evaluation of a Bed Feature to Assist in Turning and Laterally Repositioning Patients,” Hum. Factors, Vol. 58, No.

[34] R. Zall and M. R. Kangavari, “On the construction of multi-relational classifier based on canonical correlation analysis,” Int. J. Artif. Intell., Vol. 17, No. 2, pp. 23-43, 2019

Discrete Kalman Filter Invariant to Perturbations

Andrii Volovyk[1], Vasyl Kychak[2], Dmytro Havrilov[1]

[1] Department of Radio Engineering; Faculty of Infocommunications, Radio Electronics and Nanosystems; Vinnytsia National Technical University, Khmelnytske shose str., 95, 21021 Vinnytsia, Ukraine, voland@vntu.edu.ua, havrilov@vntu.edu.ua

[2] Department of Telecommunication Systems and Television; Faculty of Infocommunications, Radio Electronics and Nanosystems; Vinnytsia National Technical University, Khmelnytske shose str., 95, 21021 Vinnytsia, Ukraine, kychak.v.m@vntu.edu.ua

Abstract: Fault detection problems in dynamic objects and their localization are a very critical and rather challenging tasks for many practical applications. The Kalman-filter technology is used for these purposes most often. The correct operation indicator of the specified filter is the innovation process to be represented as a normal uncorrelated stochastic process with zero mean value and a priori calculated covariation matrix, except the specified conditions, are violated in case of unforeseen perturbations. The aim of the presented work is to develop a method allowing to restore the normal performance of the Kalman filter in the presence of uncertain disturbances. This aim is attained by applying a special one-to-one transformation of the output equation of the testing system, as a result of it, the disturbance component is modified by the extrapolation equation of the state vector dynamic system. This feature will be used in the sequel when modified Kalman filter is applied to the transformed system. The properties of the obtained filter concerning the stability of estimation errors, their convergence, and optimality are discussed. The efficiency of the method has been verified by the method of statistical modeling on a test example of a third-order dynamic system.

Keywords: states estimation; linear dynamic systems; Kalman filter; uncertain structure perturbations

1 Introduction

The fault detection problems in dynamic objects and their localization are relevant and rather difficult tasks for many practical applications [1-9]. IFAC SAFEPROCESS (Symposium on Fault Detection, Supervision and Safety for Technical Processes) defines a fault, at least, as the inadmissible deviation of one feature or parameter from its rated value to have been regulated by the standard norms [10-11]. A performance impairment can happen in separate modules of a control subject, in the regulator subsystems, switching equipment, or in observations channels, etc. The FDIR system (Fault Detection, Isolation, and Reconfiguration) is defined as a design strategy of control systems to be capable to ensure continuously functional safety or operating capacity of a control subject at beginning of a fault by its timely detection and isolation (FDI) with a possibility of the subsequent reconfiguration of the control unit in response to fault influence.

Usually, problems of fault detection and their localization are solved in two stages. On the first of them, it is necessary to make the binary decision from two mutually excluding alternative hypotheses "a system it is operational" - "the system is faulty". This stage is imperative for any functional diagnostics system. At the second stage, the place of fault emergence and its possible reason is defined. This stage, as a rule, is desirable, but isn’t regulated strictly [12]. In general, this design strategy is geared to the introduction of the redundancy concept, both by a physical layer and an analytical level.

The procedure of comparison of the duplicated signals created by the various hardware is the basis for the concept of physical redundancy introduction. For example, the same signal is observed by means of several sensors operating at different modes of operation. The standard practice implementing the hardware redundancy consists in the application of cross procedures of measuring channels cross-check, difference signals forming on the basis of a parity relations method, and further processing of the received signals by the corresponding methods, for example, using of Wavelet [41] or TP [42] transformations.

Conversely, the concept of analytical redundancy actively uses a mathematical model of a system in aggregate with the special methods of estimation considering features of the FDI systems. This concept doesn’t assume installations of the additional hardware, and in this sense is preferable in comparison with the concept of introduction of hardware redundancy. Distinctly, the maximum effect can be gained by a combination of both concepts in the uniform integrated system. However, methods of introduction of analytical redundancy are more difficult as it is necessary to guarantee stability in relation to the operating noise, unknown perturbations and incomplete information about parameters of a mathematical model.

So far the problem subject of FDI can be considered almost created. It found the reflection in the conventional classification of the existing methods, the published books, and periodical review articles. For example, the methods of analytical redundancy introduction can be separated into two major sub-classes. In the first of them the methods oriented to the application of quantitative models in an explicit form and based on are used: concepts of the parity relations [13-16]; full order observers or unknown input observers [17-20]; properties of the updated process created by a Kalman filter [21-24]; procedures of joint states estimation and unknown parameters [25-26]; stochastic algorithms [27, 29]; optimization methods [28, 30]. The general property of the above-mentioned methods is the use of specified sections of the modern control theory for the purposes to form special signals in the FDI systems. The missing information, at the same time, is generated from the results of observations.

In the second case, qualitative models on the basis of artificial intelligence methods using a mathematical apparatus of fuzzy logic [32] are applied; qualitative methods; the methods using knowledge bases; linguistic methods. For the analysis of fuzzy logic methods, we will consider a problem of difference signal forming. The difference signal, even in nominal conditions, is never equal to zero in accuracy. There is a lot of reasons for that: incomplete separation, nonlinearity, perturbations, noise, etc. Therefore, the main problem becomes to make a correct decision in the conditions of inadequate or incomplete information. As opposed to classical logic, fuzzy logic allows for making justified decisions, based on fuzzy knowledge, heuristic logic, and their combinations. Conceptually, signals processing by means of fuzzy logic consists of three stages. At the first stage, the difference signals are compared by means of special membership function. In most cases it has the triangular format. At the second stage, the smaller exit from two previous is selected. At the third stage, the procedure of center balance finding or another averaging method is used. It allows to resolve uncertainties and to lead to the probable correct decision. However, in this case, the major problem preventing of perspective technology implementation in practical applications is caused by the complexity of the training process. So, for example in [33, 34] used the basic principles of fuzzy logic for the solution of a difference signal estimation problem. The procedure of the weighed summing was made use of there instead of the categorical procedure of type "yes" - "no". In this area, it is possible to find out more about the latest advances in works [35-37] and also in the recent review publications [3, 38] related to application aspects of FDI in the context of chemical and technology objects based on AI technology.

In as much as formulation of the correct mathematical model of a control system is time-consuming method and complicated problem, many attempts to construct an acceptable qualitative diagnostic model on the basis of declarative knowledge of a system, for example, the pole analyzing of the variable, trend like "increase or decrease", a variable or a constant, etc. were made in due time. These concepts are the baseline of the qualitative method, and with their help, entirely possible, to construct the diagnostic system steadily in a sense. Moreover, comprehensive diagnostics of faults demands of, as a rule, different levels of prior information beginning from quantitative, analytical, heuristic, and finishing by expert level. It can be carried out on the basis of the expert systems functional diagnostics [39, 40] by using the complex integrated solution.

The submitted paper belongs to a subclass of the functional diagnostic methods be actively using in an explicit form quantitative mathematical model of a controlled system and relies on characteristics of the innovation process created by a Kalman filter. In [43] the possibility of faults diagnostics, using well-known statistical methods of the likelihood ratio or the generalized likelihood ratio for testing of a difference signal for "whiteness", its mean square value and an error covariation matrix of prediction was attempted for the first time. A little later, in [44, 45] was offered the adaptive estimation algorithm constructed on the basis of model conditional Kalman filter bank (MMAE-a method). The difference signal characteristics of the MMAE method were in detail studied in work [46]. Applications of these methods to FDI problems in the flight control systems are known in [47, 48].

By today, in this direction, two concepts of estimation problems with present faults and perturbations were formulated. The first of them is based on the conception of state vector expansion at the cost of connecting to it the additional unknown input associated with the active faults and perturbations influence. At the same time, it is supposed that the mathematical model of an unknown input dynamics is a priori available, and the optimal solution of an estimation problem is guaranteed by an expanded Kalman filter (EKF). However, at a large number of the considered faults and perturbations, the dimensions of the filter will be much more the control system dimensions. In [49] it was offered, by the introduction of special UV-of transformation, the procedure of EKF separation into smaller dimension constituent parts working in parallel and independently. Further, the basic idea [47] was adapted to stochastic type of faults and perturbations [51, 52]. The main efforts of researchers in this direction are made for the search of the EKF approximation methods to combine acceptable estimation accuracy with the restrictions not too complicated in terms of practical applications [53, 54].

The basis of the alternative concept is the assumption of total absence of the prior information in regard to dynamic properties of unknown inputs. In [55] it was for the first time solved this problem for the purpose of the derivation of linear unbiased estimations with the minimum generalized variance by the introduction of the certain restrictions imposed on structure of the analyzed system. In [56-57] the results [55] were generalized having applied a parametrical approach to deduce of optimum estimations. Later in [59] the optimum filter with the minimum generalized variance considering a problem of degradation characteristics inherent in the filter [55] was offered. In [60, 61] solved a fault detection problem of fault detection and their localization by means of geometrical approach, creating at the same time difference signals with the directed properties.

A specific feature above the proposed solutions is the complexity of the applied mathematical apparatus connected with the use of function spaces transformations of finite dimensions. It is to a certain extent by exposing to difficulties the practicing engineers as these sections of mathematics, often, remain outside to the standard training programs of the engineering profile specialists. Therefore, it is desirable to obtain a rather simple theoretical justification of the difference signal separation from the influence of uncertain disturbances, applying at the same time a mathematical apparatus of minimum acceptable level complexity. Unlike a traditional way of the filter structure adaptation to the mathematical model structure the authors used a reduction way of the mathematical model to the equivalent form where the disturbance component is absent in an explicit form. It allowed for it to be limited to the application of the well-known (standard) form of Kalman filter guaranteeing the derivation of the estimate state convergence in more usual terms "an bounded input – an bounded exit".

The following structure of the article is assumed: in Section 2 – the problem statement is formulated in the mathematical sense; in Section 3 – discusses a one-to-one mathematical transformation of the output of the original system, designed to absorb disturbances in the output; the results of applying the Kalman method to the transformed system are discussed in Section 4 and the main properties of the offered filter in Section 5. In Section 6 the illustrative example of the third dynamic system order for the operability purpose demonstration of the offered method is given. Subsequent sections present the results of modeling, summarize the research results, and list the literature used.

2 Problem Definition

Let's consider a linear discrete stochastic system, a mathematical model that can be described in terms of state variables

$$\mathbf{s}(k+1)=\mathbf{W}(k)\mathbf{s}(k)+\mathbf{G}(k)\mathbf{u}(k)+\mathbf{D}_s(k)\mathbf{d}(k)+\mathbf{n}_s(k); \tag{1}$$

$$\mathbf{y}(k)=\mathbf{H}(k)\mathbf{x}(k)+\mathbf{n}_y(k)\,, \tag{2}$$

where $\mathbf{s}(k)\in\square^{n}$ – the current system state, $\mathbf{y}(k)\in\square^{m}$ – the output vector, $\mathbf{u}(k)\in\square^{p}$ – exactly known control influence, $\mathbf{d}(k)\in\square^{q}$ – the indefinite structure perturbation, $\mathbf{n}_s(k)\in\square^{n}$ – the noise of a state variables, $\mathbf{n}_y(k)\in\square^{m}$ – the system output noise, $\mathbf{W}(k)$, $\mathbf{G}(k)$, $\mathbf{D}_s(k)$, $\mathbf{H}(k)$ –the known system matrices of the corresponding dimensions. The initial state $\mathbf{s}(0)\in\square^{n}$ represents a Gaussian random vector with mean value $E\{\mathbf{s}(0)\}$ and a positive definite covariation matrix $\mathbf{P}(0)$. Random sequences $\mathbf{n}_s(k)$, $\mathbf{n}_y(k)$ are independent white Gaussian noise uncorrelated with $\mathbf{s}(0)$, have zero mean values and limited covariation matrixes $E\{\mathbf{n}_s(k),\ \mathbf{n}_s^T(k)\}=\mathbf{Q}(k)$ and $E\{\mathbf{n}_y(k),\ \mathbf{n}_y^T(k)\}=\mathbf{R}(k)$, respectively. The listed assumptions coincide with those that are usually accepted in the classical theory of linear filtration without taking into account the matrix $\mathbf{D}_s(k)$ which is missing there. It is in the case under consideration supposed that perturbations $\mathbf{d}(k)$ have neither the probabilistic description nor even property of limitation from above. Otherwise stated, it is absolutely indefinite function. However, for the solvability of the problem, the following additional assumptions are introduced:

– the sequence of matrixes $\mathbf{H}(k+1)\ \mathbf{D}_{\mathrm{s}}(k)$ should be limited;

– $q \leq m$ i.e. the number of perturbations are no more than number of output sensors;

– for all $k \in \square_0$, the smallest singular values of the matrix product **H**(k + 1) $\mathbf{D}_s(k)$ not less γ, where γ – is the set of positive numbers. The last two restrictions essentially mean that the matrix product **H**(k + 1) $\mathbf{D}_s(k)$ has a full rank in the columns, and they are necessary for the perturbation absorption procedure. The task is to develop a simplified method for estimating the state vector **s**(k), free from the influence of disturbances **d**(k), based on the availability of observation results **y**(k), a sequence of precisely known control actions **u**(k) and system matrixes **W**(k), **G**(k), $\mathbf{D}_s(k)$, **H**(k). In the theory of optimal linear filtration, the stability of the Kalman filter is guaranteed by the introduction of assumptions about the controllability and observability of the system under study [3]. Similar conditions will be formulated for the case under consideration after the procedure for absorbing the disturbances has been carried out.

3 One-to-One Transformation System Exit

In this section, a local goal is pursued, namely, the justification of the procedure for absorption of the component $\mathbf{H}(k+1)\mathbf{D}_s(k)$ in the equation of state of the system by introducing a supplementary transformation of the output equation so that later it becomes possible to use the standard Kalman filter. It, in turn, will promote the forming of states estimation errors free from the influence of perturbations, subject to the transformed exit. For the first time, the problem of a difference signal separation from influence of unknown perturbations was considered in [55], where the structure of a linear filter was determined by the equation

$$\hat{\mathbf{s}}\left({}^{k+1}\!/_{k+1}\right)=\mathbf{W}(k)\hat{\mathbf{s}}\left({}^{k}\!/_{k}\right)+\mathbf{L}(k)\left[\mathbf{y}(k+1)-\mathbf{H}(k+1)\mathbf{W}(k)\hat{\mathbf{s}}\left({}^{k}\!/_{k}\right)\right]. \tag{3}$$

It is worthy of note that in this equation the component **G**(k)**u**(k) was not taken into account, since it is a precisely known quantity and is insignificant for the problem of optimal linear filter synthesis. For this case, the state vector estimation is equivalent to vector error estimation of filtering. The transfer matrix of the filter **L**(k + 1) was defined by minimization of an error covariation estimation matrix on condition that introduced restriction to be correct.

$$\mathbf{L}(k+1)\mathbf{H}(k+1)\mathbf{D}_s(k)-\mathbf{D}_s(k)=0. \tag{4}$$

This restriction guaranteed the lack of influence of a component $\mathbf{D}_s(k)\mathbf{d}(k)$ on a state estimation error. The solution of the local optimization problem taking into account (4) led to a significant complication of the process of calculating the transfer matrix **L**(k + 1) compared to the classical Kalman filter, and it was not entirely obvious how to analyze the stability of the synthesized filter. Therefore, in this article, the main attention is paid to the issue of disturbance absorption even before the filter design process and at the second stage, the modified Kalman filter option is applied to the transformed state equation (1) where the perturbation component $\mathbf{D}_s(k)\mathbf{d}(k)$ are missing.

Suppose there is some bounded matrix sequence $\mathbf{M}(k) \in \square^{n\times m}$ the specific type of which will be determined a little later. Then, relation (2) immediately implies

$$\mathbf{M}(k+1)\left[\mathbf{y}(k+1) - \mathbf{H}(k+1)\mathbf{s}(k+1) - \mathbf{n}_y(k+1)\right] = 0\,. \tag{5}$$

Further we will add to each party of the equation (5) the equation (1)

$$\begin{aligned}&\mathbf{s}(k+1) = \mathbf{W}(k)\mathbf{s}(k) + \mathbf{G}(k)\mathbf{u}(k) + \mathbf{D}_s(k)\mathbf{d}(k) + \mathbf{n}_s(k) + \\ &+ \mathbf{M}(k+1)\left[\mathbf{y}(k+1) - \mathbf{H}(k+1)\mathbf{s}(k+1) - \mathbf{n}_y(k+1)\right].\end{aligned} \tag{6}$$

After reducing such terms, we get an expression in which the equation for the output of participation no longer takes:

$$\begin{aligned}&\mathbf{s}(k+1) = \mathbf{Z}(k+1)[\mathbf{W}(k+1)\mathbf{s}(k) + \mathbf{G}(k)\mathbf{u}(k) + \mathbf{D}_s(k)\mathbf{d}(k)] + \\ &+\mathbf{M}(k+1)\mathbf{y}(k+1) + \mathbf{Z}(k+1)\mathbf{n}_s(k) - \mathbf{M}(k+1)\mathbf{n}_y(k+1),\end{aligned} \tag{7}$$

where $\mathbf{Z}(k+1) \square [\mathbf{I}_n - \mathbf{M}(k+1)\mathbf{H}(k+1)]$.

If the matrix sequence $\mathbf{M}(k + 1)$ is chosen so that in each instant of k restriction $\mathbf{Z}(k + 1)\ \mathbf{D}_s(k) = 0$ is carried out, then the equation (7) will take a form:

$$\mathbf{s}(k+1) = \mathbf{W}1(k)\mathbf{s}(k) + \mathbf{G}1(k)\mathbf{u}(k) + \mathbf{M}(k+1)\mathbf{y}(k+1) + \mathbf{w}(k), \tag{8}$$

where

$\mathbf{W}1(k) \square \mathbf{Z}(k+1)\mathbf{W}(k);\ \mathbf{G}1(k) \square \mathbf{Z}(k+1)\mathbf{G}(k);$

$\mathbf{w}(k) \square \mathbf{Z}(k+1)\mathbf{n}_s(k) - \mathbf{M}(k+1)\mathbf{n}_y(k+1)$.

In this case, the covariance matrix of the transformed state noise $\mathbf{w}(k)$ should be calculated on each computation cycle by the formula

$$\mathbf{Q}1(k) \square \mathbf{Z}(k+1)\mathbf{Q}(k)\mathbf{Z}^T(k+1)^T + \mathbf{M}(k+1)\mathbf{R}(k)\mathbf{M}^T(k+1). \tag{9}$$

Turning to equation (8) it is easy to notice that now the disturbing effect $\mathbf{D}_s(k)\ \mathbf{d}(k)$ is excluded from further transformations in an explicit form. At this stage, it is important that the perturbation is absorbed at the moment $k + 1$ instead of k an instant is important. The condition that the transmission matrix of the filter under consideration must satisfy $\mathbf{M}(k + 1)$ similar to that introduced in [12], namely $\mathbf{Z}(k + 1)\mathbf{D}_s(k) = [\mathbf{I}_n - \mathbf{M}(k + 1)\mathbf{H}(k + 1)]\mathbf{D}_s(k) = 0$. However, there is a significant difference here. Expression in square brackets provides more degrees of freedom in choosing the value of the matrix transmission coefficient $\mathbf{M}(k + 1)$ since it is not related to solving the minimization problem. This matrix can be determined by solving the matrix equation $\mathbf{M}(k+1)\mathbf{H}(k+1)\mathbf{D}_s(k) = \mathbf{D}_s(k)$. Since it is assumed that the assumption is valid that in the matrix product $\mathbf{H}(k+1)\mathbf{D}_s(k)$ the number of columns does not exceed the number of rows, this means that the solution for $\mathbf{M}(k+1)$ must exist. In most practical applications, this inequality is satisfied. Taking this remark into account, we obtain [5]

$$\mathbf{M}(k+1) = \mathbf{D}_s(k)\left[\mathbf{H}(k+1)\mathbf{D}_s(k)\right]^{*}, \tag{10}$$

where the symbol $[\cdot]^*$ the pseudoinverse Moore-Penrose matrix is denoted [63].

4 Result of the Kalman Filter Method Application

Returning to the transformed state model (8), it is easy to see that state vectors $\mathbf{s}(k)$, an entrance $\mathbf{u}(k)$, and an exit $\mathbf{y}(k)$ remained the same. Thus, the original model (1) and the modified (8) are essentially equivalent, since they describe the same system, and the component $\mathbf{M}(k+1)\ \mathbf{y}(k+1)$ can be interpreted as a new known input. In this case, there are no formal obstacles to the application of the standard estimation procedure by the Kalman method, since the disturbance component $\mathbf{D}_s(k)\ \mathbf{d}(k)$ is not present. Then, the application of the classical Kalman filter to the transformed model (8)

$$\mathbf{y}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{n}_y(k),$$

$$\mathbf{s}(k+1) = \mathbf{W}1(k)\mathbf{s}(k) + \mathbf{G}1(k)\mathbf{u}(k) + \mathbf{M}(k+1)\mathbf{y}(k+1) + \mathbf{w}(k),$$

will generate in a recurrent form the following estimates of the states $\mathbf{s}^*(k/k)$ and their covariation matrixes $\mathbf{P}(k/k)$ for all $k \in \square_0$,

$$\mathbf{P}\left({}^{k+1}\!/_{k}\right) = \mathbf{W}1(k)\mathbf{P}\left({}^{k}\!/_{k}\right)\mathbf{W}1^T(k) + \mathbf{Q}1(k); \tag{11}$$

$$\mathbf{K}(k+1) = \mathbf{P}\left({}^{k+1}\!/_{k}\right)\mathbf{H}^T(k+1)\mathbf{P}^{-1}{}_{\mathbf{r}}(k+1); \tag{12}$$

$$\mathbf{P}_{\mathbf{r}}(k+1) \square E\left\{\mathbf{r}(k+1)\mathbf{r}^T(k+1)\right\} = \mathbf{H}(k+1)\mathbf{P}\left({}^{k}\!/_{k}\right)\mathbf{H}^T(k+1) + \mathbf{R}(k+1); \tag{13}$$

$$\mathbf{P}\left({}^{k+1}\!/_{k+1}\right) = \left[\mathbf{I}_n - \mathbf{K}(k+1)\mathbf{H}(k+1)\right]\mathbf{P}\left({}^{k+1}\!/_{k}\right); \tag{14}$$

$$\mathbf{r}(k+1) = \mathbf{y}(k+1) - \mathbf{H}(k+1)\mathbf{s}^*\left({}^{k+1}\!/_{k}\right); \tag{15}$$

$$\mathbf{s}^*\left({}^{k+1}\!/_{k}\right) = \mathbf{W}1(k)\mathbf{s}^*\left({}^{k}\!/_{k}\right) + \mathbf{G}1(k)\mathbf{u}(k) + \mathbf{M}(k+1)\mathbf{y}(k); \tag{16}$$

$$\mathbf{s}^*\left({}^{k+1}\!/_{k+1}\right) = \mathbf{s}^*\left({}^{k+1}\!/_{k}\right) + \mathbf{K}(k+1)\mathbf{r}(k+1). \tag{17}$$

$$\begin{aligned} &\mathbf{s}^*(\%) = E\{\mathbf{s}(0)\}, \\ &\mathbf{P}(\%) = \mathbf{P}(0). \end{aligned} \tag{18}$$

where matrixes $\mathbf{Z}(k + 1)$, $\mathbf{M}(k + 1)$, $\mathbf{W}1(k)$, $\mathbf{G}1(k)$ have to be previously calculated according to formulas (7), (10), (8), respectively, with initial conditions (18).

At the same time, it is necessary to emphasize that the offered filter differs from a normal Kalman filter a little. First, the distribution matrix of again entered control input is calculated (10) using a pseudoinverse concept, and secondly, more significantly, the covariation matrix of the generalized perturbation, being non-stationary, has to be updated in each computing cycle. In these repeated calculations there is the absorption perturbations procedure with indefinite structure in the implicit form. Besides, the seeming simplicity of the offered filter presents the additional problem in the form of generalized perturbation correlation components. Usually in practice, for simplicity of the estimation procedure neglect this correlation often. At the same time, the filter becomes quasi-optimal filter.

5 Analysis of the Modified Kalman Filter Properties

In order for the modified Kalman filter to be guaranteed to be stable, it is necessary to formulate two more constraints, in addition to those already given in the second section. Their essence is reduced to uniform full observability of the pair $[\mathbf{W}1(k), \mathbf{H}(k)]$ and uniform full controllability of the pair $\{\mathbf{W}1(k)[\mathbf{I}_n - \mathbf{K}(k)\mathbf{H}(k)]\mathbf{Q}^{1/2}(k)\}$, defined using the Gram matrix. They differ somewhat from those accepted in the theory of linear Kalman filtration since this theory cannot be directly applied to the specific problem under consideration. Moving on to the discussion of the properties of the synthesized filter, the following should be noted:

1) Data limitation. If the above-introduced assumptions are valid, then the recursively calculated matrices $\mathbf{P}(k + 1/k + 1)$ and $\mathbf{P}(k + 1/k)$, so and $\mathbf{P}_r(k + 1)$, $\mathbf{K}(k)$ are limited. In other words, this property guarantees the boundedness of all recurrent computations, excluding the estimates of the state vector. Due to the presence of white Gaussian noise, the state estimation errors, in principle, cannot be limited (11), (15) but the second point is limited – the covariance matrix of filtering errors. This property is important for applications where all calculations must be performed in real-time.

2) Stability. At the made assumptions, dynamic errors of state estimates (17), (18) will only be exponentially stable, since in this case one of the main conditions of the Kalman filter theory is violated – the reversibility of the transition matrix of states $\mathbf{W}1(k)$. In fact, this matrix according to expression (8) is always singular.

However, if we introduce the notation for the estimation error of the state vector $\mathbf{s}^{\sim}(k/k) \triangleq \mathbf{s}^{\sim}(k) - \mathbf{s}^{*}(k/k)$ then the dynamics of estimation errors can be represented by the equation

$$\begin{aligned}\tilde{\mathbf{s}}\left({}^{k+1}\!/_{k+1}\right) &= \left[\mathbf{I}_n - \mathbf{K}(k+1)\mathbf{H}(k+1)\right]\mathbf{W}1(k)\tilde{\mathbf{s}}\left({}^{k}\!/_{k}\right) + \\ &+ \left[\mathbf{I}_n - \mathbf{K}(k+1)\mathbf{H}(k+1)\right]\mathbf{w}(k) - \mathbf{K}(k+1)\mathbf{n}_y(k+1).\end{aligned} \tag{19}$$

It is common knowledge that the convergence of estimation errors is defined only by the determined member of equation (19) which is characterized by the expression $\mathbf{W}1(k)[\mathbf{I}_n - \mathbf{K}(k+1)\mathbf{H}(k+1)]$. As calculations $\mathbf{K}(k+1)$ are determined, then the stability of the dynamics of filtering errors is provided only by the assumptions made with respect to $\mathbf{W}1(k)$, $\mathbf{H}(k+1)$, $\mathbf{Q}1(k)$. Therefore, these properties don't affect the correlation processes $\mathbf{w}(k)$ and $\mathbf{n}_y(k)$. More complete proof of the estimations convergence can be deduced using of Lyapunov functions or (and) having investigated at the same time stability of the Riccati solutions at a singular transfer matrix [60]. However, it goes beyond the objects set in this work.

3) Optimality. It should be noted here that ignoring the correlation between processes $\mathbf{w}(k)$, $\mathbf{n}_y(k)$ leads to the loss of optimality of the modified filter. However, as shown by further modeling, these losses are relatively small. Besides, it is not entirely obvious how one should take into account the correlation caused by the introduced transformation of the equation of the system's output, and this can be the subject of further research.

6 Results of the Method Feasibility Testing

As the illustrative example, untied any specific application, let's consider the continuous third-order dynamic system with transfer function [64]

$$F(p) \triangleq \frac{U_{out}(p)}{U_{in}(p)} = \frac{1}{\left(T_1^2 p^2 + 2\xi T_1 p + 1\right)\left(T_2 p + 1\right)}$$

where matrixes T_1, T_2 – the time constants of the oscillatory and aperiodic links connected consistently, ξ – the damping factor ($1 < \xi < 0$), $p = (\sigma + j\omega)$ – the complex variable. As numerical values of the specified parameters we will choose $T_1 = 1$, $T_2 = 4$, $\xi = 0{,}5$. Let's convert this system in terms of a state variable. For this purpose we will use the method of direct programming [65] and will perform the following transformations:

$$U_{out}(p) = F(p)U_{in}(p) = \frac{1}{\left(T_1^2 p^2 + 2\xi T_1 p + 1\right)\left(T_2 p + 1\right)} U_{in}(p) =$$

$$=\frac{p^{-3}U_{in}(p)}{p^{-3}+(2\xi T_1+T_2)p^{-2}+\left(T_1^2+2\xi T_1T_2\right)p^{-1}+T_1^2T_2}.$$

Let's enter new denotation

$$E(p)\square\frac{U_{in}(p)}{p^{-3}+(2\xi T_1+T_2)p^{-2}+\left(T_1^2+2\xi T_1T_2\right)p^{-1}+T_1^2T_2}.$$

We will solve the obtained relation relatively $E(p)$

$$E(p)\square-\frac{1}{T_1^2T_2}E(p)p^{-3}-\frac{(2\xi T_1+T_2)}{T_1^2T_2}E(p)p^{-2}-\frac{\left(T_1^2+2\xi T_1T_2\right)}{T_1^2T_2}E(p)p^{-1}+\frac{1}{T_1^2T}U_{in}(p).$$

As state variables we will choose the integrators outputs and will make a functional diagram having the next parameters:

$$\frac{\left(T_1^2+2\xi T_1T_2\right)}{T_1^2T_2}\square a;\quad\frac{(2\xi T_1+T_2)}{T_1^2T_2}\square b;\quad\frac{1}{T_1^2T_2}\square c.$$

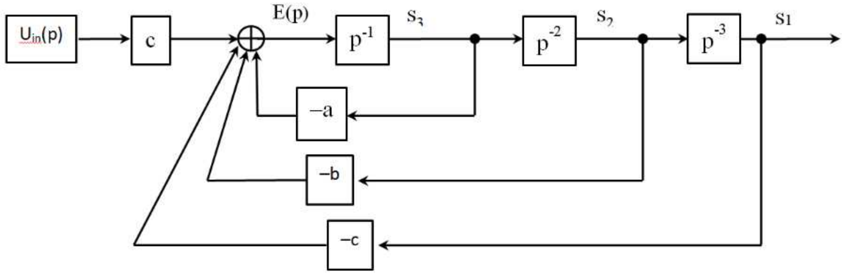


Figure 1

The function circuit of an illustrative system in terms of the state variables

On the basis of the function circuit we make the equation system

$$\begin{cases}\dot{s}_1(t)=\ \ 0\cdot s_1(t)+1\cdot s_2(t)+0\cdot s_3(t)+0\cdot u_{\textit{вх}}(t);\\ \dot{s}_2(t)=\ \ 0\cdot s_1(t)+0\cdot s_2(t)+1\cdot s_3(t)+0\cdot u_{\textit{вх}}(t);\\ \dot{s}_3(t)=-c\cdot s_1(t)-b\cdot s_2(t)-a\cdot s_3(t)+c\cdot u_{\textit{вх}}(t).\end{cases}$$

Therefore, system matrixes in the absence of perturbations will be such:

$$\mathbf{s}(t)=\begin{bmatrix}s_1(t)\\ s_2(t)\\ s_3(t)\end{bmatrix};\ \ \mathbf{W}(t)=\begin{bmatrix}0&1&0\\0&0&1\\-c&-b&-a\end{bmatrix};\ \mathbf{G}(t)=\begin{bmatrix}0\\0\\c\end{bmatrix};\ \mathbf{H}(t)=\begin{bmatrix}1&0&0\\0&0&1\end{bmatrix}.$$

Under determining an observation matrix it was supposed that the structure of measuring means allows a possibility of states measurements $\mathbf{s}_1(t)$ and $\mathbf{s}_3(t)$.

The matrix of perturbations distribution was defined from the following conditions [66]. It was supposed that a perturbations source are changes of parameters *a, b, c*, and changes of matrixes **W**(*t*), **G**(*t*) we will provide linear approximations Δ**W**(*t*), Δ**G**(*t*), i.e

$$\mathbf{D}_s(t)\mathbf{d}(t)=\mathbf{D}_s(t)\left\{\left[\Delta a(t)\ \Delta b(t)\ \Delta c(t)\right]\mathbf{s}(t)+\Delta c(t)u_{in}(t)\right\};$$

- when determining a matrix $\mathbf{D}^T_s$ (*t*)=[1 0 0]^T we will be limited to a case when changes of parameters influence a component $\mathbf{s}_1(t)$.

The system discrete equivalent was calculated by formulas [63]:

$$\mathbf{W}(k)=\mathbf{e}^{\mathbf{W}\Delta T};\quad \mathbf{G}(k)=\mathbf{W}^{-1}\left[\mathbf{W}(k)-\mathbf{I}_3\right]\mathbf{G};\quad \mathbf{D}_s(k)=\mathbf{W}^{-1}\left[\mathbf{W}(k)-\mathbf{I}_3\right]\mathbf{D}_s$$

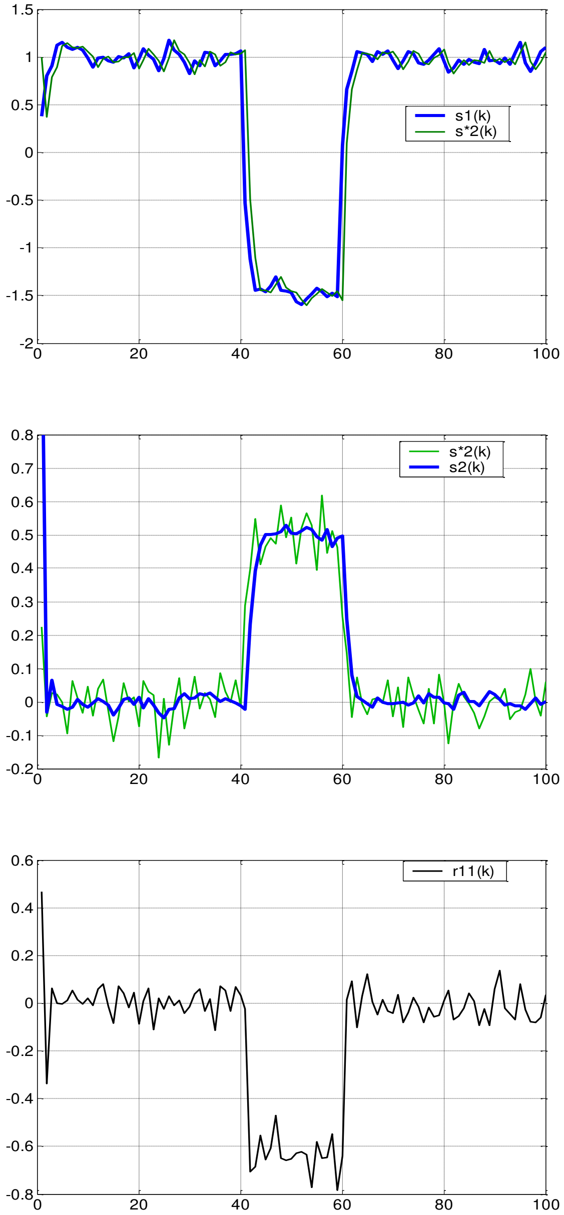
At above preset values T_1, T_2, ξ and ΔT = 4 system matrixes accept values

$$\mathbf{W}(k)=\begin{bmatrix}0.4729 & 0.5765 & 0.6260\\ -0.1565 & -0.3096 & -0.2060\\ 0.0515 & 0.1010 & -0.0521\end{bmatrix};\ \mathbf{G}(k)=\begin{bmatrix}0.5271\\ 0.1565\\ -0.0515\end{bmatrix};\ \mathbf{D}_s(k)=\begin{bmatrix}3.2121\\ -0.5271\\ -0.1565]\end{bmatrix}$$

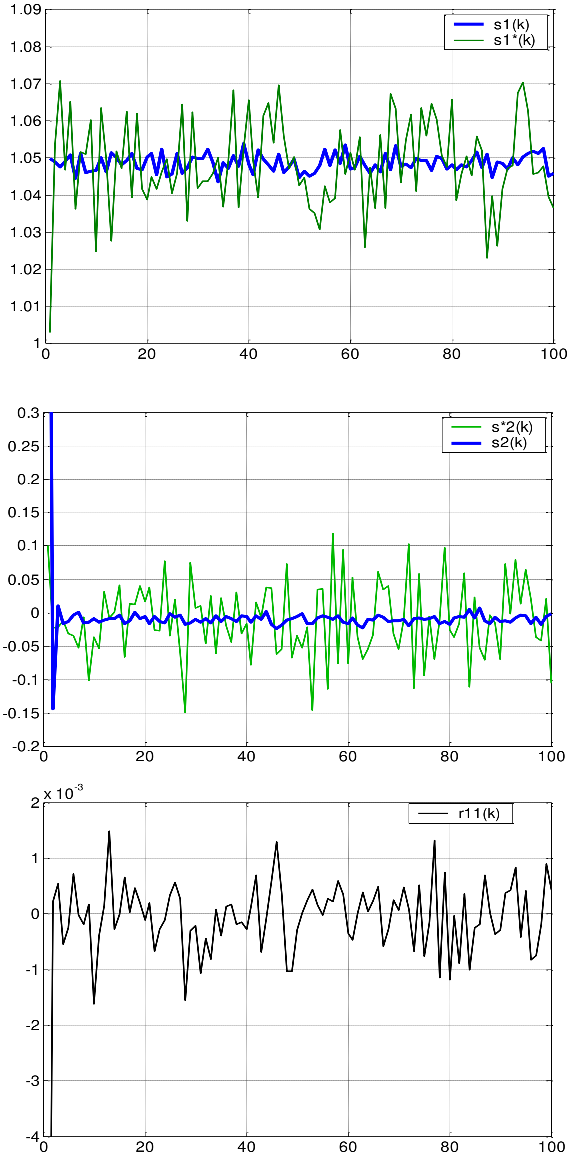
Covariance matrices of noise processes $\mathbf{n}_s(k)$, $\mathbf{n}_y(k)$, took the values $\mathbf{Q} = 0.0025\mathbf{I}_{(3)}$; $\mathbf{R} = 0.0001\mathbf{I}_{(2)}$, respectively, and the mean values were assumed to be zero. The modeling process was carried out in the computing MATLAB environment according to formulas (11)-(18). The perturbations process was imitated by expression:

$$d(k)=\begin{cases}k\le 40, & d(k)=0;\\ k>40, & d(k)=\ -0.2+\ 0.025*randn;\\ k\ge 60, & d(k)=0.\end{cases}$$

The following values were taken as the initial conditions: $\mathbf{s}^T(0) = [1; 0; -1]^T$ $\mathbf{u}(k) = 1$; $\mathbf{s}^{T*}(0/0) = [0; 0; 0]^T$; $\mathbf{P}(0/0) = \mathbf{I}_{(3)}$. Results of estimating the first two components of the state vector $\mathbf{s}_1^*(k/k)$, $\mathbf{s}_2^*(k/k)$ for system transient response and differential signal **r**(*k*) are shown in Fig. 2 a, b.


1.5
1
0.5
0
-0.5
-1
-1.5
-2
0
20
40
60
80
100
s1(k)
s*2(k)
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
0
20
40
60
80
100
s*2(k)
s2(k)
0.6
0.4
0.2
0
-0.2
-0.4
-0.6
-0.8
0
20
40
60
80
100
r11(k)

a)



b)

Figure 2

Modeling results of states estimation process: a) standard Kalman filter; b) modified Kalman filter

If the Kalman filter works correctly, then the difference process, called the updating process, is an uncorrelated Gaussian random process with a zero mean value and a covariance matrix recursively calculated by the formula [46]:

$$\mathbf{P}_{\mathbf{r}}\left(k+1\right)\square\ E\left\{\mathbf{r}\left(k+1\right)\mathbf{r}^{T}\left(k+1\right)\right\}=\mathbf{H}\left(k+1\right)\mathbf{P}\left({}^{k}\!/\!_{k}\right)\mathbf{H}^{T}\left(k+1\right)+\mathbf{R}\left(k+1\right). \tag{20}$$

Matching charts **r**(*k*) located in the bottom line of Fig. 2, confirms the correct operation of only the modified Kalman filter, the estimates error are not affected by the appeared disturbance. Consequently, the problem of decoupling (decomposition) of the estimation process from disturbances with an indefinite structure is satisfactorily solved.

Conclusions

The problem of obtaining estimates of the states of linear dynamic systems, free from the influence of disturbances, the structure of which is not defined, is relevant for many of the applied research, including in the field of functional diagnostics.

Unlike the traditional way based on an adaptation of the filter structure to the model set structure of, authors solved a problem by modification of the model set model to an equivalent form where the perturbations component in an explicit form is absent. It allowed for being limited to the application of one of Kalman filter forms guaranteeing convergence of states estimations.

In comparison with other known design methods, the proposed method is natural and simple, and the conditions for the existence of a solution are easily verified. Considerations are presented regarding the guarantee of the properties of the obtained solution: stability, boundedness and optimality. Since the issues under consideration have not yet found sufficient coverage in the periodicals, there is reason to believe that the presented results introduce an element of novelty into the research topics related to obtaining estimates of the dynamical systems states that are indifferent to disturbances of an indefinite structure.

References

[1] V. Venkatasubramanian, R. Rengaswamy, and S. N. Kavuri, “A review of process fault detection and diagnosis: Part I: Quantitative model-based methods,” Computers & Chemical Engineering, Vol. 27, No. 3, pp. 293-311, Mar. 2003, doi: 10.1016/S0098-1354(02)00160-6

[2] V. Venkatasubramanian, R. Rengaswamy, and S. N. Kavuri, “A review of process fault detection and diagnosis: Part II: Qualitative models and search strategies,” Computers & Chemical Engineering, Vol. 27, No. 3, pp. 313-326, Mar. 2003, doi: 10.1016/s0098-1354(02)00161-8

[3] V. Venkatasubramanian, R. Rengaswamy, S. N. Kavuri, and K. Yin, “A review of process fault detection and diagnosis: Part III: Process history based methods,” Computers & Chemical Engineering, Vol. 27, No. 3, pp. 327-346, Mar. 2003, doi: 10.1016/s0098-1354(02)00162-x

[4] H. Long and X. M. Wang, “Aircraft Fuel System Diagnostic Fault Detection Through Expert System”, 2008 7th World Congress on Intelligent

Control and Automation, 2008, pp. 7104-7107, doi: 10.1109/WCICA.2008.4594020

[5] K. S. Gaeid and H. A. F. Mohamed, “Diagnosis and Fault Tolerant Control of the Induction Motors Techniques a Review”, Australian Journal of Basic and Applied Sciences, 4(2), 2010, pp. 227-246

[6] J. L. Aravena and F. N. Chowdhury, "Fault detection of flight critical systems," 20th DASC. 20th Digital Avionics Systems Conference (Cat. No.01CH37219), 2001, pp. 1C6/1-1C6/10 Vol. 1, doi: 10.1109/DASC.2001.963315

[7] C. Hajiyev, F. Caliskan, “Fault diagnosis and reconfiguration in flight control systems”, Springer Science & Business Media 2003, 342 p. doi: 10.1007/978-1-4419-9166-9

[8] F. Kimmich and R. Isermann, “Model based fault detection for the injection, combustion and engine-transmission,” IFAC Proceedings Volumes, Vol. 35, No. 1, pp. 203-208, 2002, doi: 10.3182/20020721-6-es-1901.00765

[9] L. Salvatore, B. Pires, M. C. M Campos and M. B. A. De Souza Jr M. B., “A Hybrid Approach to Fault Detection and Diagnosis in a Diesel Fuel Hydrotreatment Process”, Proceedings of world academy of science, engineering and technology, Vol. 7 August, 2005, pp. 379-384

[10] D. van Schrick, “Remarks on Terminology in the Field of Supervision, Fault Detection and Diagnosis,” IFAC Proceedings Volumes, Vol. 30, No. 18, pp. 959-964, Aug. 1997, doi: 10.1016/s1474-6670(17)42524-9

[11] R. Isermann, Fault-Diagnosis Systems. Springer Berlin Heidelberg, 2006, doi: 10.1007/3-540-30368-5

[12] M. Blanke, M. Kinnaert, J. Lunze, and M. Staroswiecki, Diagnosis and Fault-Tolerant Control. Springer Berlin Heidelberg, 2016, doi: 10.1007/978-3-662-47943-8

[13] R. Isermann, “Model-based fault-detection and diagnosis – status and applications,” Annual Reviews in Control, Vol. 29, No. 1, pp. 71-85, Jan. 2005, doi: 10.1016/j.arcontrol.2004.12.002

[14] J. J. Gertler and R. Monajemy, “Generating directional residuals with dynamic parity relations,” Automatica, Vol. 31, No. 4, pp. 627-635, Apr. 1995, doi: 10.1016/0005-1098(95)98494-q

[15] J. J. Gertler, X. Fang, and Q. Luo, “Detection and Diagnosis of Plant Failures: The Orthogonal Parity Equation Approach,” in Control and Dynamic Systems, Elsevier, 1990, pp. 159-216, doi: 10.1016/b978-0-12-012737-5.50010-4

[16] R. J. Patton and J. Chen, “A Review of Parity Space Approaches to Fault Diagnosis,” IFAC Proceedings Volumes, Vol. 24, No. 6, pp. 65-81, Sep. 1991, doi: 10.1016/s1474-6670(17)51124-6

[17] Frank P. M., “On-line fault detection in uncertain nonlinear systems using diagnostic observers“: a survey. Int. J. Systems Sci., Vol. 25/12, pp. 2129-2154, 1994

[18] V. Krishnaswami and G. Rizzoni, “A Survey of Observer Based Residual Generation for FDI,” IFAC Proceedings Volumes, Vol. 27, No. 5, pp. 35-40, Jun. 1994, doi: 10.1016/s1474-6670(17)48000-1

[19] J. Chen, H. Zhang, “Robust detection of faulty actuators via unknown input observers”, Proceedings of International Journal of Systems Science, Vol. 22, Iss. 10, 1991, pp. 1829-1839, doi: 10.1080/00207729108910753

[20] A. Volovyk, V. Kychak, D. Kudriavtsev, D. Havrilov, A. Yarovyi and L. Krylik, “Simultaneous Estimation in Linear Dynamic Systems with the Indeterminate Structure Disturbances”, 2020 IEEE 40th International Conference on Electronics and Nanotechnology (ELNANO), Kyiv, Ukraine, 2020, pp. 651-655, doi: 10.1109/ELNANO50318.2020.9088884

[21] C. Yang, J. Zheng, X. Ren, W. Yang, H. Shi and L. Shi, "Multi-Sensor Kalman Filtering with Intermittent Measurements" in IEEE Transactions on Automatic Control, Vol. 63, No. 3, pp. 797-804, March 2018, doi: 10.1109/TAC.2017.2734643

[22] P. Bania and J. Baranowski, "Bayesian estimator of a faulty state: Logarithmic odds approach," 2017 22nd International Conference on Methods and Models in Automation and Robotics (MMAR), 2017, pp. 253-257, doi: 10.1109/MMAR.2017.8046834

[23] J. Keller and D. D. J. Sauter, "Kalman Filter for Discrete-Time Stochastic Linear Systems Subject to Intermittent Unknown Inputs," in IEEE Transactions on Automatic Control, Vol. 58, No. 7, pp. 1882-1887, July 2013, doi: 10.1109/TAC.2013.2264739

[24] S. Huang, K. K. Tan and T. H. Lee, "Fault Diagnosis and Fault-Tolerant Control in Linear Drives Using the Kalman Filter," in IEEE Transactions on Industrial Electronics, Vol. 59, No. 11, pp. 4285-4292, Nov. 2012, doi: 10.1109/TIE.2012.2185011

[25] S. Simani, C. Fantuzzi, and R. J. Patton, "Model-based fault diagnosis in dynamic systems using identification techniques", Advances in Industrial Control. Springer-Verlag London, 2003, p. 282

[26] R. Isermann, “Supervision, fault-detection and fault-diagnosis methods — An introduction,” Control Engineering Practice, Vol. 5, No. 5, pp. 639-652, May 1997, doi: 10.1016/s0967-0661(97)00046-4

[27] H. Wang and W. Lin, “Applying observer based FDI techniques to detect faults in dynamic and bounded stochastic distributions,” International Journal of Control, Vol. 73, No. 15, pp. 1424-1436, Jan. 2000, doi: 10.1080/002071700445433

[28] Zall R., Kangavari M., "On the construction of multi-relational classifier based on canonical correlation analysis ", International Journal of Artificial Intelligence, Vol. 17, No. 2, pp. 23-43, 2019

[29] A. Y. Volovik, L. V. Krylik, I. M. Kobylyanska, A. Kotyra, S. Amirgaliyeva, “Methods of stochastic diagnostic type observers”, Proc. SPIE 10808, Photonics Applications in Astronomy, Communications, Industry, and High-Energy Physics Experiments 2018, 108082X (1 October 2018), doi: 10.1117/12.2501693

[30] J.-Y. Keller, “A Fault Detection Filter Including an Adaptive Noise Cancellation Strategy,” European Journal of Control, Vol. 13, No. 6, pp. 627-638, Jan. 2007, doi: 10.3166/ejc.13.627-638

[31] J. Stoustrup and H. H. Niemann, “Fault estimation – a standard problem approach,” Int. J. Robust Nonlinear Control, Vol. 12, No. 8, pp. 649-673, 2002, doi: 10.1002/rnc.716

[32] L. A. Zadeh, K. S. Fu, K. Tanaka, M. Shimura, and C. V. Negoita, “Fuzzy Sets and Their Applications to Cognition and Decision Processes,” IEEE Trans. Syst., Man, Cybern., Vol. 7, No. 2, Feb. 1977, doi: 10.1109/tsmc.1977.4309670

[33] P. M. Frank and N. Kiupel, “Fuzzy supervision and application to lean production†,” International Journal of Systems Science, Vol. 24, No. 10, pp. 1935-1944, Oct. 1993, doi: 10.1080/00207729308949605

[34] P. M. Frank, “Application of Fuzzy Logic to Process Supervision and Fault Diagnosis,” IFAC Proceedings Volumes, Vol. 27, No. 5, pp. 507-514, Jun. 1994, doi: 10.1016/s1474-6670(17)48077-3

[35] W. Cheng and C. Juang, "A Fuzzy Model With Online Incremental SVM and Margin-Selective Gradient Descent Learning for Classification Problems," in IEEE Transactions on Fuzzy Systems, Vol. 22, No. 2, pp. 324-337, April 2014, doi: 10.1109/TFUZZ.2013.2254492

[36] C.-F. Juang, Y.-Y. Lin, and R.-B. Huang, “Dynamic system modeling using a recurrent interval-valued fuzzy neural network and its hardware implementation,” Fuzzy Sets and Systems, Vol. 179, No. 1, pp. 83-99, Sep. 2011, doi: 10.1016/j.fss.2011.05.015

[37] R. Precup, T. Teban, A. Albu, A. Borlea, I. A. Zamfirache and E. M. Petriu, "Evolving Fuzzy Models for Prosthetic Hand Myoelectric-Based Control," in IEEE Transactions on Instrumentation and Measurement, Vol. 69, No. 7, pp. 4625-4636, July 2020, doi: 10.1109/TIM.2020.2983531

[38] R. J. Patton, F. J. Uppal, and C. J. Lopez-toribio, “Soft Computing Approaches to Fault Diagnosis for Dynamic Systems: A Survey,” IFAC Proceedings Volumes, Vol. 33, No. 11, pp. 303-315, Jun. 2000, doi: 10.1016/s1474-6670(17)37377-9

[39] R. Isermann, “Integration of Fault Detection and Diagnosis Methods,” IFAC Proceedings Volumes, Vol. 27, No. 5, pp. 575-590, Jun. 1994, doi: 10.1016/s1474-6670(17)48088-8

[40] Claudiu Pozna, Radu-Emil Precup, Applications of signatures to expert systems modeling. Acta Polytechnica Hungarica 11(2): January 2014, p. 21-39

[41] Merry R. J. E.: Wavelet Theory and Applications. Eindhoven, 2005, p. 49

[42] E. Hedrea, R. Precup, R. Roman, and E. M. Petriu, “Tensor product- based model transformation approach to tower crane systems modeling,” Asian J Control, Vol. 23, No. 3, pp. 1313-1323, Mar. 2021, doi: 10.1002/asjc.2494

[43] R. K. Mehra and J. Peschon, "An innovations approach to fault detection and disgnosis in dynamic systems". Automatica, Vol. 7, pp. 637-640, 1971

[44] D. Magill, "Optimal adaptive estimation of sampled stochastic processes," in IEEE Transactions on Automatic Control, Vol. 10, No. 4, pp. 434-439, October 1965, doi: 10.1109/TAC.1965.1098191

[45] D. Lainiotis, "Optimal adaptive estimation: Structure and parameter adaption," in IEEE Transactions on Automatic Control, Vol. 16, No. 2, pp. 160-170, April 1971, doi: 10.1109/TAC.1971.1099684

[46] P. D. Hanlon and P. S. Maybeck, "Characterization of Kalman filter residuals in the presence of mismodeling," in IEEE Transactions on Aerospace and Electronic Systems, Vol. 36, No. 1, pp. 114-131, Jan. 2000, doi: 10.1109/7.826316

[47] Hajiyev Ch. M., "Fault detection in multidimensional systems based on statistical analysis of Kalman filter". IFAC Symposium on Fault Detection Supervision and Safety for Technical Processes, SAFERPROCESS – Helsinki, 1994, Vol. 1, pp. 45-49

[48] C. Hajiyev and F. Caliskan, “Sensor/actuator fault diagnosis based on statistical analysis of innovation sequence and Robust Kalman Filtering,” Aerospace Science and Technology, Vol. 4, No. 6, pp. 415-422, Sep. 2000, doi: 10.1016/s1270-9638(00)00143-7

[49] B. Friedland, "Treatment of bias in recursive filtering," in IEEE Transactions on Automatic Control, Vol. 14, No. 4, pp. 359-367, August 1969, doi: 10.1109/TAC.1969.1099223

[50] J. Chen and R. J. Patton, “Optimal filtering and robust fault diagnosis of stochastic systems with unknown disturbances,” IEE Proceedings - Control

Theory and Applications, Vol. 143, No. 1, pp. 31-36, Jan. 1996, doi: 10.1049/ip-cta:19960059

[51] M. Ignagni, "Optimal and suboptimal separate-bias Kalman estimators for a stochastic bias," in IEEE Transactions on Automatic Control, Vol. 45, No. 3, pp. 547-551, March 2000, doi: 10.1109/9.847741

[52] Fu-Chuang Chen and Chien-Shu Hsieh, "Optimal multistage Kalman estimators," in IEEE Transactions on Automatic Control, Vol. 45, No. 11, pp. 2182-2188, Nov. 2000, doi: 10.1109/9.887678

[53] Chien-Shu Hsieh, "Extension of the robust two-stage Kalman filtering for systems with unknown inputs," TENCON 2007 - 2007 IEEE Region 10 Conference, 2007, pp. 1-4, doi: 10.1109/TENCON.2007.4429133

[54] C. Hsieh, "A unified extension of the robust two-stage Kalman filter and its application to functional filtering," 2009 American Control Conference, 2009, pp. 3824-3829, doi: 10.1109/ACC.2009.5159992

[55] P. K. Kitanidis, "Unbiased minimum-variance linear state estimation," Automatica, Vol. 23, No. 6, pp. 775-778, Nov. 1987, doi: 10.1016/0005-1098(87)90037-9

[56] J. Y. Keller and M. Darouach, "Two-stage Kalman estimator with unknown exogenous inputs," Automatica, Vol. 35, No. 2, pp. 339-342, Feb. 1999, doi: 10.1016/s0005-1098(98)00194-0

[57] A. Volovyk, V. M. Kychak, "Detection Filter Method in Diagnostic Problems for Linear Dynamic Systems", Visnyk NTUU KPI Seriia – Radiotekhnika Radioaparatobuduvannia, 2021, Iss. 84, pp. 30-39, DOI: https://doi.org/10.20535/RADAP.2021.84

[58] S. Gillijns and B. De Moor, "Unbiased minimum-variance input and state estimation for linear discrete-time systems with direct feedthrough," Automatica, Vol. 43, No. 5, pp. 934-937, May 2007, doi: 10.1016/j.automatica.2006.11.016

[59] Chien-Shu Hsieh, "Optimal Minimum-Variance Filtering for Systems with Unknown Inputs," presented at the 2006 6th World Congress on Intelligent Control and Automation, 2006, doi: 10.1109/wcica.2006.1712679

[60] M. A. Massoumnia, "A geometric approach to the synthesis of failure detection filters," in IEEE Transactions on Automatic Control, Vol. 31, No. 9, pp. 839-846, September 1986, doi: 10.1109/TAC.1986.1104419

[61] J. White and J. Speyer, "Detection filter design: Spectral theory and algorithms," in IEEE Transactions on Automatic Control, Vol. 32, No. 7, pp. 593-603, July 1987, doi: 10.1109/TAC.1987.1104682

[62] C. de Souza, M. Gevers and G. Goodwin, "Riccati equations in optimal filtering of nonstabilizable systems having singular state transition

matrices," in IEEE Transactions on Automatic Control, Vol. 31, No. 9, pp. 831-838, September 1986, doi: 10.1109/TAC.1986.1104415

[63] Arthur Albert, Regression and the Moore-Penrose pseudoinverse, New York : Academic Press, 180 p., 1972

[64] Makarov I. M., Mensky B. M. Linear automatic systems (elements of theory, calculation methods and reference material), Mashinostroenie, 1982, 505 p.

[65] Julius T. Tou, Modern Control Theory, NY, McGraw-Hill, 1964, 427 p.

[66] A. Varga, Solving Fault Diagnosis Problems. Springer International Publishing. Vol. 84. 2017, p. 394, doi: 10.1007/978-3-319-51559-5

Development and Design Optimisation of a Small Floating Hydroelectric Power Plant

Tomas Kalina[1], Ladislav Illes[2], Martin Jurkovic[1*], Vladimir Luptak[3]

[1]The Faculty of Operation and Economics of Transport and Communications, University of Zilina, Univerzitna 1, 01026 Zilina, Slovakia; tomas.kalina@fpedas.uniza.sk, martin.jurkovic@fpedas.uniza.sk

[2]MULTI Engineering Services, Dunajske nabrezie 4726, 94501 Komarno, Slovakia; ladislav.illes@multi.engineering

[3]Department of Transport and Logistics, Faculty of Technology, Institute of Technology and Business in Ceske Budejovice, 37001 Ceske Budejovice, Czech Republic; luptak@mail.vstecb.cz

Abstract: This study deals with the research and development of the optimal design of a small floating hydroelectric power plant by theoretical analysis and the subsequent conceptual design of the optimal variant. A computational fluid dynamics (CFD) system is used for theoretical analyses of flow, flow around, free surface properties, and motion of bodies in the water. The aim of this study is to identify the optimal geometry and construction of a small floating hydroelectric power plant. In the study, five different versions of floating pontoons are designed and analysed in the first phase. CFD analysis is used to determine the choice of the most suitable concept, which is further modified based on the calculation results. The result of the study is the design of a suitable design solution, which obviously achieves higher efficiency compared to a conventional water wheel. Finally, the further direction of research is presented, with a focus on maximising the performance and further optimisation of the small floating hydroelectric power plant structures.

Keywords: optimisation; power plant; computational fluid dynamics (CFD); water wheel; hydraulic power

1 Introduction

Hydropower plants are powerful tools for reducing greenhouse gas emissions. There are several perspectives on the efficiency of hydropower plants and on the process of building them in order to generate electricity from renewable sources [1]. For this purpose, large waterworks are being built, including hydropower plants, which, however, are not currently operating efficiently and it is, therefore, necessary to modernise and rebuild them to operate more efficiently [2].

In addition, the impact of hydropower plants on the ecosystem is negative [3]. Another approach is to build small hydropower plants, which are often also questioned for several reasons [4]. The most significant problem in building small hydropower plants is the major intervention in the natural river flow. This is most often achieved by completely damming the flow. This maximises the potential of the flow, but has several negative effects. Such negative effects can only be eliminated by a partial damming of the stream, which makes it possible to preserve, at least partially, the natural state of the riverbed. Small hydropower plants can therefore cause serious damage to the river ecosystem. When comparing small and large hydropower plants, small hydropower plants have a greater negative impact per megawatt of electricity produced [5, 6].

For the above reasons, trends in the construction of conventional hydropower plants should be directed towards modernisation and streamlining of the work of large hydropower plants rather than the massive construction of small hydropower plants. Small hydropower plants should be designed primarily for those areas where the population density is sparse and where it is not efficient to draw electricity through power lines [7, 8].

These claims are also confirmed by a study carried out in Norway, which compared the environmental impact of small and large hydropower plants [9]. However, deciding on the direction of development is often not just a question of assessing the effectiveness or environmental impact, but also of other circumstances, such as political priorities. However, at present, the attenuation of the proactive construction of small hydropower plants can be seen, which can be considered a good signal to stabilise the situation [10, 11]. The main environmental consequences of hydropower plants can be classified by their impact on the aquatic ecosystem [12].

By damming the stream, the natural flow in the riverbed is disrupted, which results in degradation of the bottom and banks or erosion. Dams disrupt the flow, worsen the quality of water, block the movement of animals and sediments in the river, destroy fish habitats and prevent the natural migration of fish, with even built fish paths not alleviating the issue [13].

However, there are other methods than those mentioned above that make it possible to obtain electricity from renewable sources. In our study, we search for alternatives to building small hydropower plants. One method is to install small floating devices. Such devices are suitable for rivers with small slopes and low flow velocities, with research in this area already under way. An example is the floating hydroelectric power plant on the Mura River in Hungary. This power plant achieves an electrical output of 5-10 kW. We focus on the possibility of increasing the efficiency of floating equipment by various methods of technical improvement [14].

We present the research and development of the optimal construction of a small floating hydroelectric power plant by theoretical analysis and subsequent conceptual design of the most suitable variant. As a computational fluid dynamics (CFD) software tool, the Autodesk Simulation CFD 2016 system is used in this project for theoretical analyses of fluid flow, flow around, free surface properties, and body motions in water. The provided CFD simulations allowed us to model the dynamics of fluid flow in configurations that are calculated on the base of the Navier–Stokes equations implemented in the software and solving the numerical problem using the finite volume method [15].

Meaning of words *Optimization* and *Optimum* in this paper: in general they stand for a process and its result the goal of which is to find the best technical idea and solution giving the highest possible performance of the plant but also fulfilling several different requirements, e.g. issues concerning the river flow conditions, the river bank infrastructure, the possible environmental impacts, the protection of the wheels, the complexity of manufacturing the plant, the economic profitability, etc.

So, they are not used in a purely mathematical meaning, rather in technical.

The aim of this study is to identify the optimal geometry and construction of a small floating hydroelectric power plant intended for the Slovak section of the Danube and the lower part of the Váh River. The following partial tasks are solved using a series of theoretical calculations, analyses, and simulations:

- Optimal shape of the water wheel and its size;
- Optimal shape of the water wheel blades;
- Optimal number of blades;
- Optimal location of the hydropower plant with respect to the riverbank and flow depth;
- Optimal mutual placement of several water wheels by comparing the effectiveness of different layouts;
- Optimal shape of the water corridor and water supply to the water wheels;
- Use of water flow barriers and rectifiers to increase the efficiency of the water wheel;
- Possibility of using the lower boundary of the water corridor;
- Optimal water wheel rotation speed.

The following general procedure was used to solve the individual sub-tasks:

- Drawing up various alternatives;
- Theoretical comparison of the effectiveness of the various alternatives;
- Identification of two or several of the most efficient alternatives based on theoretical calculations.

2 Theoretical Background

Based on the nature of the discussed flows, the floating power plant under investigation is a free surface, hydrokinetic type that uses purely the kinetic energy of water flow. Classic water wheels are the most suitable for such electricity production in the difficult conditions of our rivers, especially when the economic aspect is also considered.

A similar single-wheel floating hydroelectric power plant installed on the Mura River in Hungary achieves an average overall efficiency of 55% [14]. This object was considered a model at the beginning of the development and a higher resulting efficiency level was set as a target. When determining the input parameters of the tunnel flow and the initial dimensions of the wheels, the possible main dimensions of the double-wheel bearing floating pontoon were also considered and their parameters are given in Table 1.

Table 1

Main dimensions of a double-float pontoon

L [m]	B [m]	H [m]	T [m]
max. 40	11.4	3.0	2.0

where:

- L [m] is the length of a double-float pontoon;
- B [m] is the breadth of a double-float pontoon;
- H [m] is the height of a double-float pontoon;
- T [m] is the draught of a double-float pontoon.

The rectangular cross-sectional area of the tunnel is calculated by:

$$S = b \cdot t\ [m^2]\,, \tag{1}$$

where:

- $b\ [m]$ is the width of the rectangular section of the tunnel;
- $t\ [m]$ is the depth of the rectangular section of the tunnel.

The volume flow rate is calculated by:

$$Q = S \cdot v\ [m^3 \cdot s^{-1}]\,, \tag{2}$$

where:

- $v\ [m \cdot s^{-1}]$ is the flow velocity.

The mass flow rate is calculated by:

$$q = \rho \cdot Q \; [kg \cdot s^{-1}], \tag{3}$$

where:

- $\rho \; [kg \cdot m^{-3}]$ is the density of water.

The maximum theoretical power of the water flow kinetic energy is calculated by:

$$P_W = \frac{1}{2} \cdot q \cdot v^2 \; [kW]. \tag{4}$$

The hydraulic power of the power plant is determined by:

$$P_H = P_W \cdot \eta_H \; [kW], \tag{5}$$

where:

- $\eta_H [-]$ is the hydraulic efficiency of the wheel and tunnel.

The mechanical power of the power plant is determined by:

$$P_M = P_H \cdot \eta_M \; [kW], \tag{6}$$

where:

- η_M [-] is the mechanical efficiency of the bearings and gears.

The delivered electric power of the power plant is determined by:

$$P_E = P_M \cdot \eta_E \; [kW], \tag{7}$$

where:

- $\eta_E \; [-]$ is the electrical efficiency of the generator.

The overall efficiency of the hydropower plant is calculated by:

$$\eta_T = \eta_H \cdot \eta_M \cdot \eta_E \; [-]. \tag{8}$$

The equation used to estimate the output hydraulic power of a single hydrokinetic water wheel is:

$$P_O = \frac{1}{2} \cdot \rho \cdot S \cdot C_p \cdot v^3 \; [kW], \tag{9}$$

where:

- $C_p \; [-]$ is the power coefficient.

However, determining the power coefficient C_p is very difficult, and very different approaches are reported in the literature. In contrast, the use of a single wheel has been excluded during the development and this equation only has a very limited application for multiple wheels that interact with each other.

In this development task, the hydraulic power was determined from the output values of the CFD analyses, which were the torques and angular velocities of the water wheels, by:

$$P_H = M \cdot \omega \ [kW] , \tag{10}$$

where:

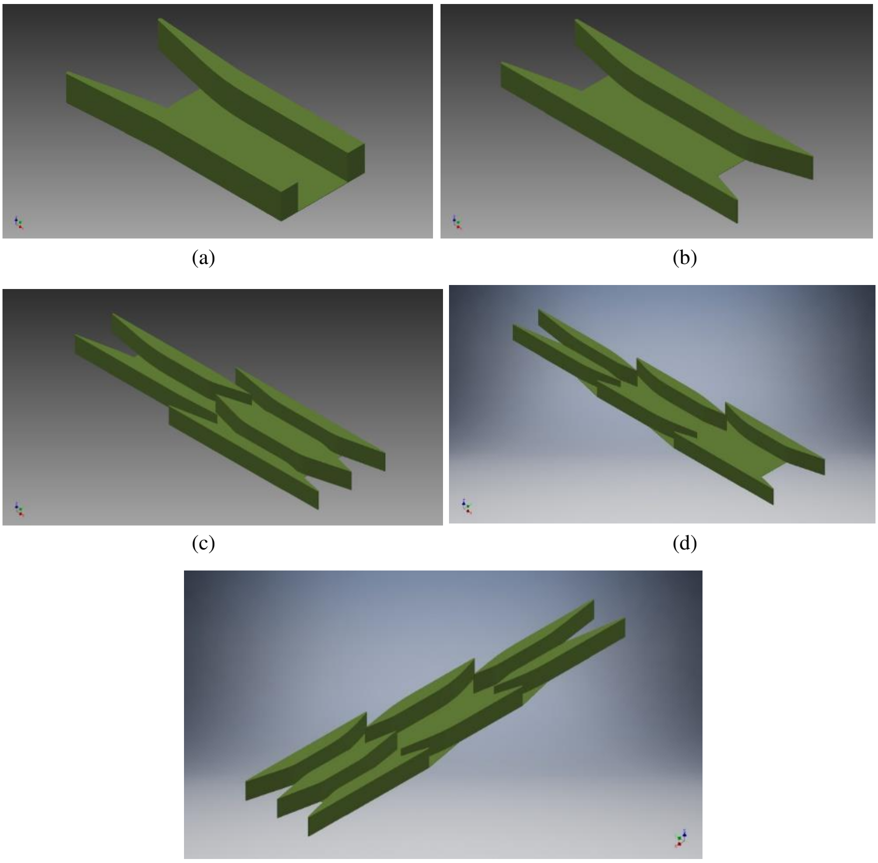
- M$[Nm]$ is the hydrodynamic torque;
- $\omega[rad \cdot s^{-1}]$ is the angular velocity.

From the hydraulic power value obtained, the hydraulic efficiency η_H was calculated using Eq. (5). Other parameters, such as mechanical and electric power, can only be determined when the corresponding efficiencies are known. The efficiencies depend on the choice of other components of the power plant, on the bearing of the wheel shaft, on the transmission devices and on the electric current generator.

3 Methods

This study consists of the analysis of different configurations. The most common method that allows for computer modeling of flow is CFD. The principle and limitations of this method in relation to flow modeling have been addressed in several studies [16-18]. The results of modeling, as well as the time of their implementation, depend mainly on the number of elements (and nodes) and also on the quality of computer technology [19]. Computer-aided design software is used for the two- and three-dimensional design of floating power plant elements [20].

Five different versions of floating pontoons, from A to E (Figure 1), were designed and analysed by means of computer-aided design and CFD software (see example CFD results in Figures 1f and 3c). The flow velocity properties inside the tunnels were considered in order to eliminate unfavourable shapes with high resistances (losses). Step-by-step, the pontoon versions with the greatest drag forces, the greatest flow decelerations, and the smallest mass flow (flux) between pontoons were eliminated. For comparable values, "simpler" shapes with fewer water wheels have always had the advantage.



(a) (b)

(c) (d)

(e)

Figure 1

Different versions of the pontoons: (a) pontoon A, basic version of two floats/two wheels; (b) pontoon B, improved version of two floats/two wheels; (c) pontoon C, five floats/three wheels; (d) pontoon D, six floats/three wheels; (e) pontoon E, seven floats/four wheels

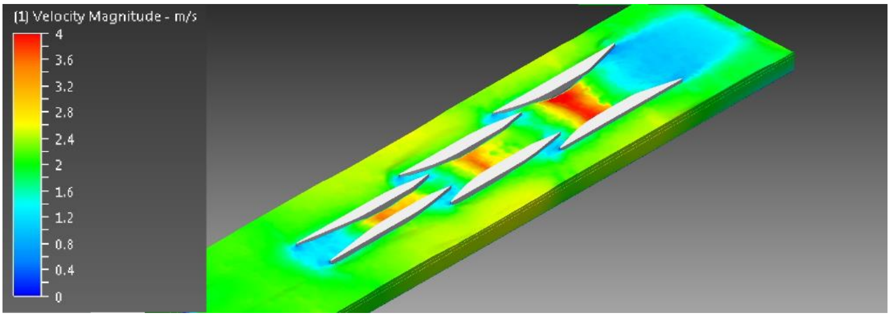


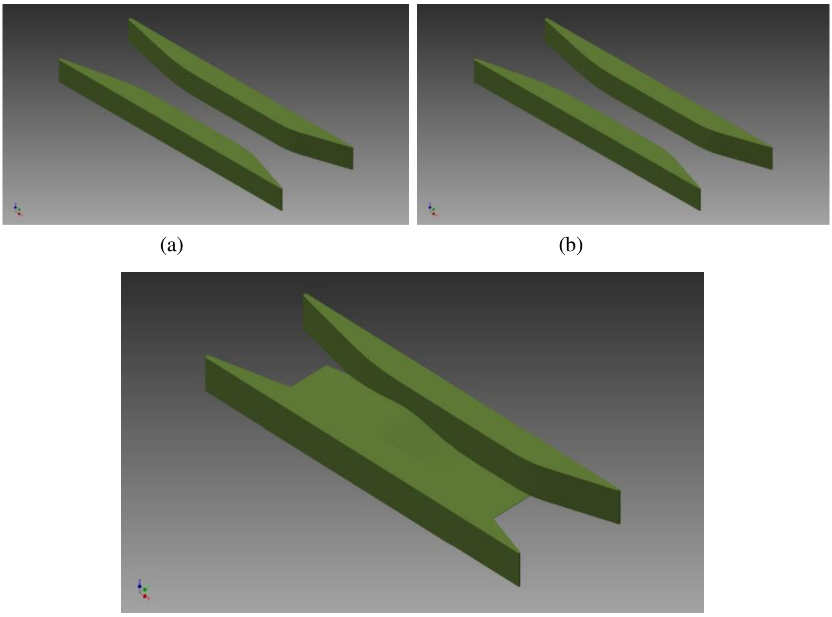
Figure 2

Example CFD output showing free surface flow velocities

All preliminarily designed versions of the pontoons passed an optimisation CFD test, where three possible variations were examined:

- Pontoon open from below in between the floats;
- Pontoon from below closed by a flat bottom between floats;
- Pontoon closed with a specially shaped inner bottom.

Flow analysis was performed for all versions from A to E. Figure 3 shows pontoon B in more detail.



(a) (b)

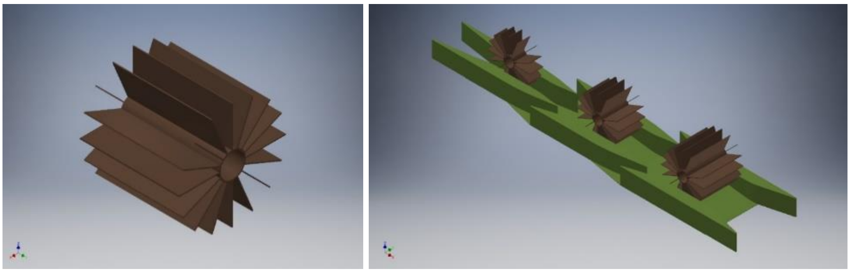
(c)

Figure 3

An example flow analysis for pontoon B: (a) pontoon open at the bottom between the floats; (b) pontoon closed from below by a flat bottom; (c) pontoon closed with the shaped inner bottom surface

CFD analysis shows that the most suitable design is the pontoon provided by a flat bottom between the floats. During the consideration also other factors have been taken into account like the influence of the river bed on the induced turbulences or the protection of the rotating parts against floating objects, etc.

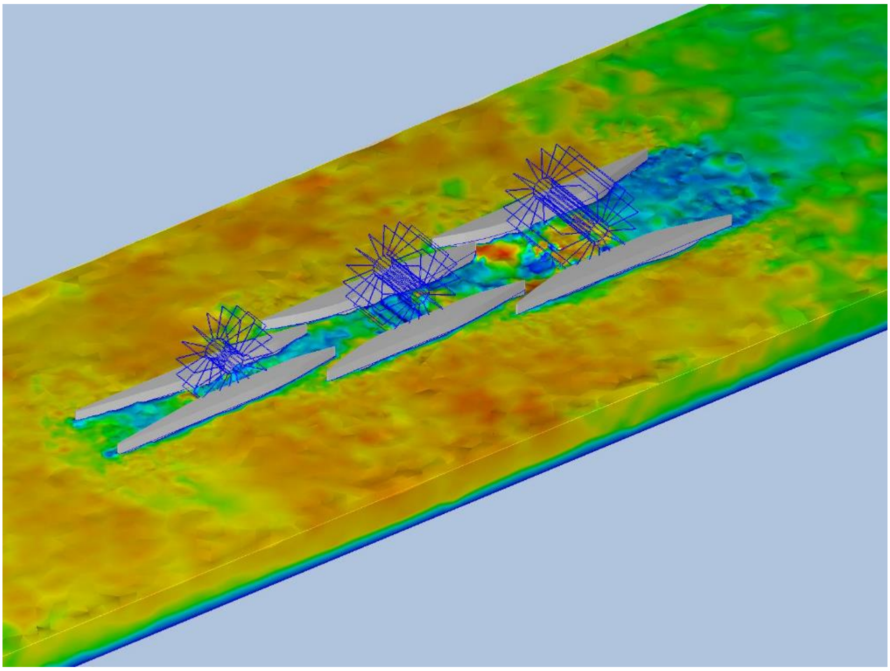
In the next step, the interaction of water wheels (basic version) with individual floating pontoons was tested by a simplified CFD analysis. The analysis was performed in a relatively short time for all versions from A to E, with pontoon D presented in Figure 4.



(a) (b)

Figure 4

Interaction of water wheels with pontoon D. (a) Three-dimensional model of the basic version of the wheel. (b) Three-dimensional model of the pontoon/wheel assembly



(c)

Figure 5

Example CFD output showing free surface flow pattern

As rotating water wheels have appeared in the CFD domain to work together with the stationary pontoons, the most important selection criteria became the generated hydrodynamic torque and the angular velocity of the wheels. Based on the results of these CFD analyses, it was decided that further analysis and development would be limited to only pontoons B and D. The calculations also showed that the third row of wheels no longer had significant performance advantages.

As on this basis and also from an economic point of view, the optimal number of wheels was set at two. As a result, version D was modified to a two-tier assembly and is now labelled as pontoon F.

The computational domain and the solver of the Simulation CFD software were configured as follows:

- transient task with free water surface (multiphase),
- k-ε turbulence model,
- tetrahedral mesh, mean size 50 mm in refined near areas, 250 mm in far areas,
- rotating regions surrounding the wheels.

Grid independence calculations particularly for this project have not been performed. Based on several previously analysed similar configurations it can be stated that the expected difference in resulting values would be in range 1-2%, if the analysis was performed for a refined mesh with half-size elements. This is acceptable, the difference is less than the numerical error of the process.

The aim of the analysis was not to obtain exact absolute values of hydrodynamic quantities, but rather to confront relative values and optimize by comparison methods.

4 Results

The computational tunnel is a purely theoretical case of a channel open from above, having a rectangular cross-section of a sectional area identical with the cross-section of a solid filled wheel. In other words, the wheel completely fills the cross-section or is enclosed by the boundary surfaces of the computational volume (CFD domain). Such a domain is suitable for comparative analyses because the impact of the environmental factors is minimised.

4.1 Water Wheel Optimisation and Analysis of Hydraulic Performance in Computational Tunnel

Three optimisation steps were performed sequentially based on the comparative calculations and analyses. The size of the blades was determined by the depth of the water and the dimensions of the wheels determined the main dimensions of the pontoon. The number of blades was gradually increased until the maximum hydraulic power of the wheel was reached. The following versions of the water wheel were proposed as outputs:

Version 1: Optimisation of the basic version of the wheel, with curved blades, used instead of straight ones (Figure 6). The radius of curvature and the number of blades were optimised based on the resulting hydrodynamic properties of the wheel.

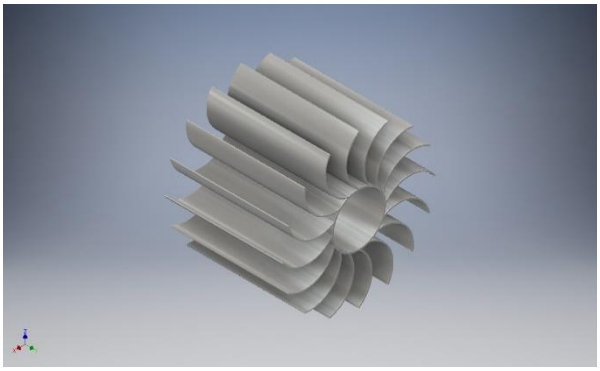


Figure 6
Water wheel version 1 with curved blades

Version 2: Results of the second step of the wheel optimisation. In order to achieve a smoother course of hydraulic pressures and torques, but also to reduce vibrations and possible resonant states of the wheel pairs, the wheel was divided into three mutually twisted "discs" (Figure 7).

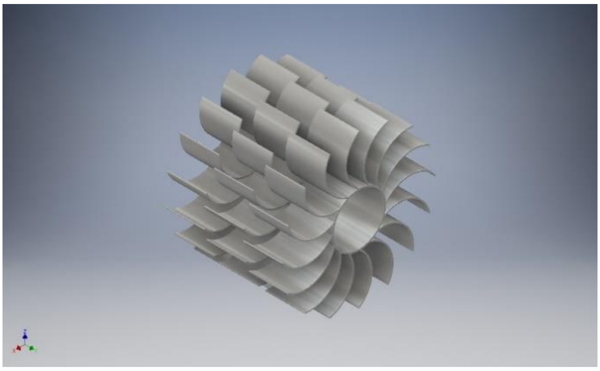


Figure 7
Water wheel version 2 with three mutually twisted discs

Version 3: The third step of the wheel optimisation. In order to eliminate significant axial flows of water between the blades and reduce the related losses, the individual discs were separated by thin circular plates (Figure 8).

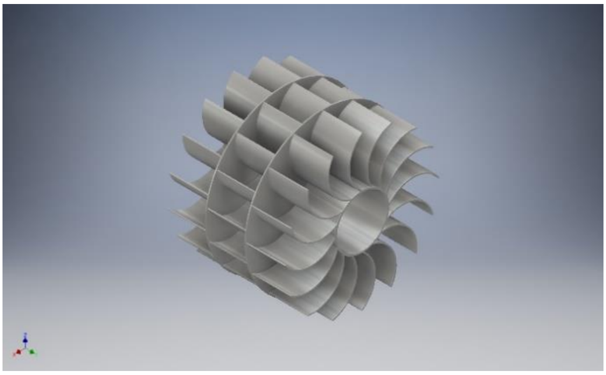
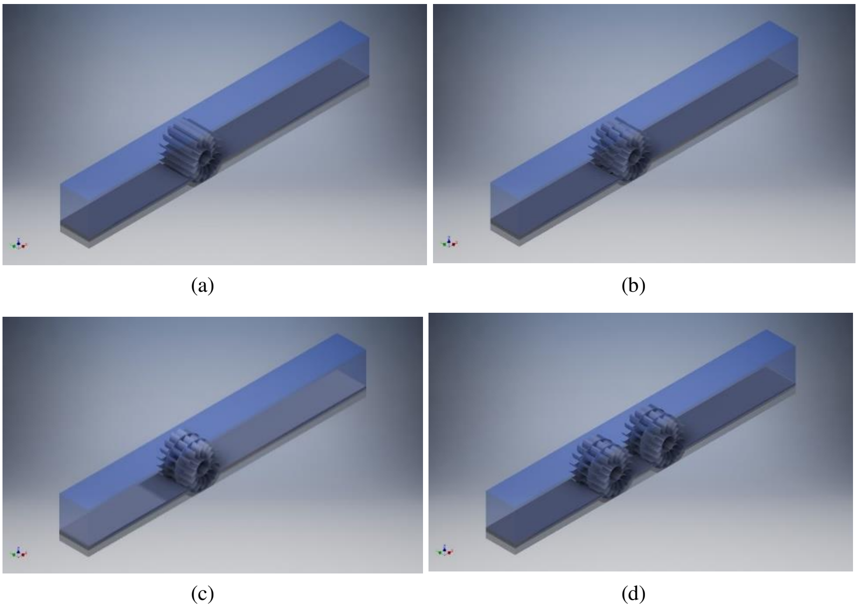


Figure 8
Water wheel version 3 with three twisted discs separated by thin plates

4.2 Analysis of Hydraulic Power in the Computational Tunnel

CFD simulations of the individual versions were performed for the same calculation tunnel with the same settings for water flow velocities of 2, 2.5, 3 and 3.5 m/s (Figure 9).



(a)

(b)

(c)

(d)

Figure 9
CFD analyses of the three different kinds of blades in the tunnel: (a) wheel version 1; (b) wheel version 2; (c) wheel version 3; (d) two wheels of version 3

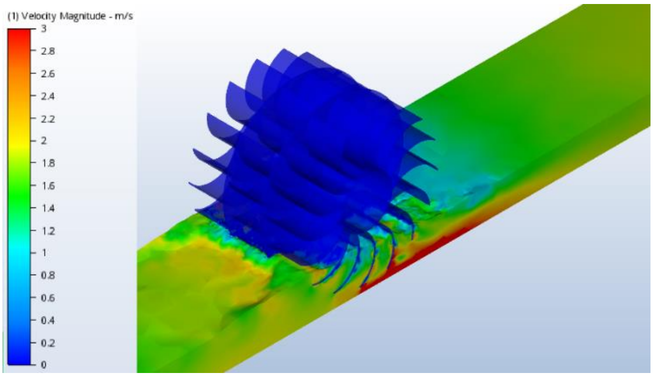


Figure 10
An example CFD output of flow velocities in a tunnel

4.3 Geometrical Design of Floating Pontoons and Synergy Analysis

The geometry of only two of the following types of pontoons was optimised, which were selected for further development in the previous optimisation steps (the flat bottom of the pontoons is not shown in the next figures).

Pontoon B: The floating pontoon of type B (Figure 11a) is a double-float construction, closed from below with a flat bottom. It is suitable for the installation of two identical water wheels in a row. It has one water inlet at the front and one outlet at the rear.

Pontoon F: Pontoon type F (Figure 11b) was created as a combination of pontoons B and D. It has four floats, is closed from below with a flat bottom, and is suitable for installing two water wheels of different widths in a row. It has one water inlet at the front, two side inlets, and one outlet at the rear.

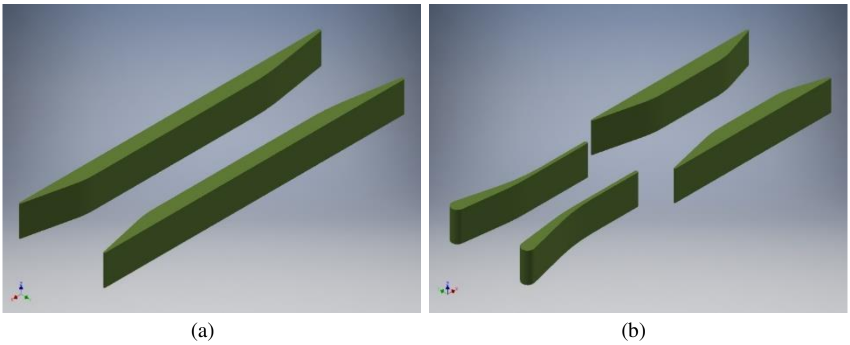
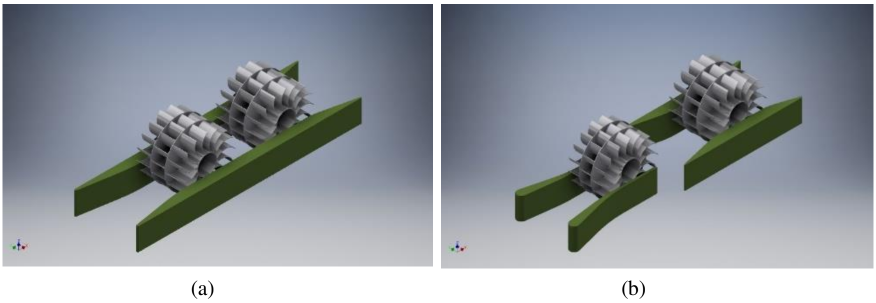

(a) (b)

Figure 11
Computation models of pontoons (a) B and (b) F

All in-depth CFD analyses of hydraulic performance were performed uniformly for a water flow velocity of 2 m/s. The aim of the calculations was to compare the performances and efficiencies of the two most promising arrangements of pontoons B and F. Based on the previous results and considerations, two-tier-type arrangements were selected and equipped with pairs of version 3 water wheels (Figure 12). In version F, the front wheel was made narrower to improve the lateral water supply to the rear wheel.


(a)

(b)

Figure 12

CFD models of different pontoons with version 3 wheels for pontoons (a) B and (b) F

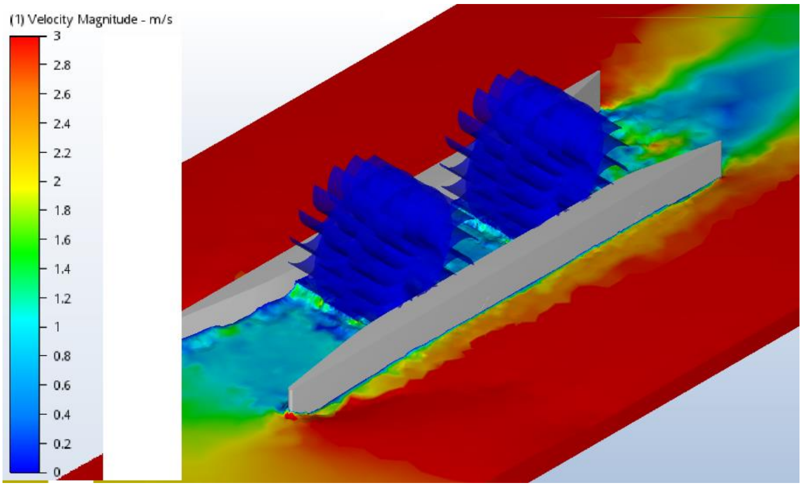


Figure 13

An example CFD output for flow velocities in the pontoon area

5 Discussion

Based on the CFD analyses performed in the process of water wheel optimisation, a gradual increase in hydraulic performance was observed, the course of which is shown in Figure 14. A uniform water flow velocity of 2 m/s was used in all calculations.

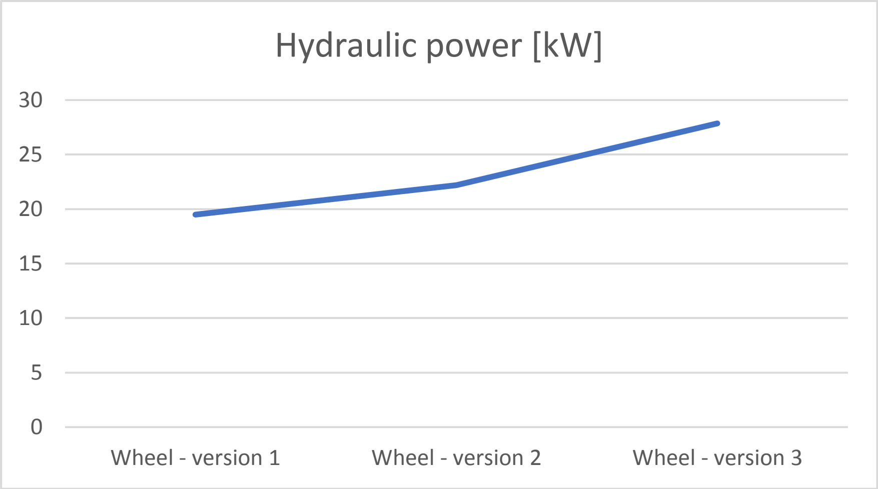


Figure 10

Increase of hydraulic power depending on degree of wheel optimisation

Figure 14 shows that wheel version 2 has an increased hydraulic power compared to wheel version 1 by 16% and to the wheel version 3 by up to 46%.

CFD analyses were also performed to determine the dependency of hydraulic power on the velocity of water flow (Figure 15). Simplified CFD analyses were performed in a theoretical calculation tunnel for a single wheel assembly with the most efficient geometry of wheel version 3.

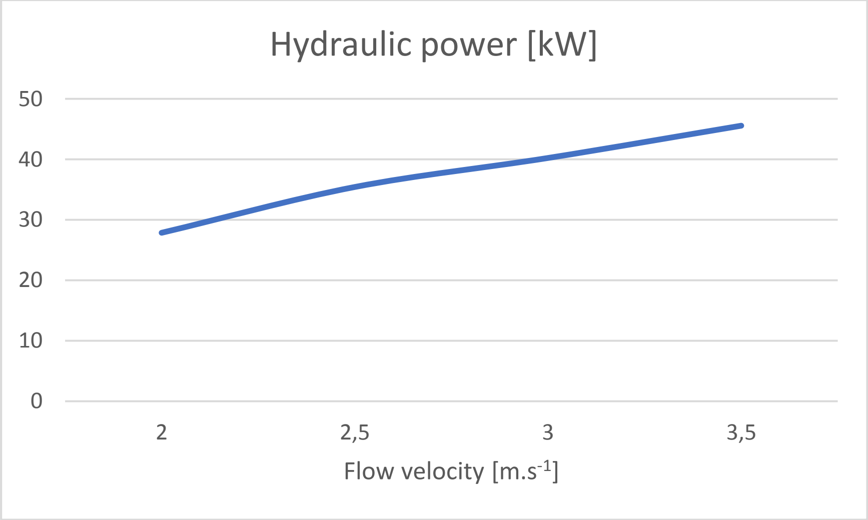


Figure 15
Dependence of hydraulic power on flow velocity

From Figure 15, it is clear that by increasing the flow velocity from the base value (2 m/s), the increase of hydraulic power does not follow closely the increase of kinetic energy of the water flow, i.e., the hydraulic efficiency of the assembly decreases. Such a reduction, especially at higher speeds, could also be affected by the applied calculation method (simplified domain); however, these results had no effect on the further development process. For the final evaluation, more accurate analyses were performed and velocities higher than 2.5 m/s were not examined (they are not realistic at the chosen geographical locations).

Based on the results of the in-depth CFD analyses of the plant assembly, the highest hydraulic power and thus the best hydraulic efficiency was achieved by pontoon F with a pair of version 3 wheels. The average value of the calculated hydraulic power reached P_H = 59.5 kW. Analyses were performed for a flow velocity of 2 m/s. This value was chosen as the characteristic velocity realistically achievable in the target locations of the concerned sections of the Danube and the Váh River.

The total hydraulic efficiency of a small floating power plant was determined from the dimensional parameters of the wheel/pontoon system and from the physical properties of the water flow using Eqs. (1) to (5).

For input values of P_H = 59.5 kW, b = 10.0 m, t = 2.0 m, ρ = 1000.0 kg/m^3 and v = 2 m/s, the physical quantities listed in Table 2 were determined successively and the resulting total hydraulic efficiency is 0.744.

Table 2

Calculation parameters and hydraulic efficiency of power plant

S [m^2]	**Q [m^3/s]**	**q [kg/s]**	**P_W [kW]**	**P_H [kW]**	**η_H [-]**
20.00	40.0	40000.00	80.0	59.5	0.744

where:

- S [m^2] is the cross-sectional area of the tunnel;
- Q [m^3/s] is the volume flow rate;
- q [kg/s] is the mass flow rate;
- P_W [kW] is the theoretical power of the water flow;
- P_H [kW] is the hydraulic power of the power plant;
- η_H [-] is the total hydraulic efficiency.

Based on the total hydraulic efficiency obtained and the estimated values of mechanical and electrical efficiency, the electric power and overall efficiency of the power plant were further determined by means of Eqs. (6) to (8).

The chosen values for the theoretical efficiencies were:

- η_M = 0.96 - this mechanical efficiency is achievable with a direct generator drive;
- η_E = 0.95 - most of the current generators achieve such electrical efficiency by default.

The calculation of the usable theoretical energy of water flow and the electric power of the power plant was performed for the flow velocity range of 1.75-2.50 m/s, assuming that the hydraulic efficiency does not change significantly in the calculation interval. The results are shown in Table 3 and the plots are presented in Figure 16.

Table 3

Calculation parameters and theoretical electric power of power plant

v [m/s]	**Q [m^3/s]**	**q [kg/s]**	**P_W [kW]**	**P_H [kW]**	**P_E [kW]**
1.75	35.0	35000.00	53.6	39.9	36.4
2.00	40.0	40000.00	80.0	59.5	54.3
2.25	45.0	45000.00	113.9	84.8	77.3
2.50	50.0	50000.00	156.3	116.3	106.0

where:

- v [m/s] is the flow velocity;
- Q [m^3/s] is the volume flow rate;

- q [kg/s] is the mass flow rate;
- P_W [kW] is the theoretical power of the water flow;
- P_H [kW] is the hydraulic power of the power plant;
- P_E [kW] is the delivered electric power of the floating plant.

The previous analyses and considerations result in a value of 0.68 for the total efficiency η_T, while based on Figure 15 (single-stage arrangement in a computational tunnel with a simplified analysis), it is very likely that the hydraulic efficiency will decrease at higher speeds.

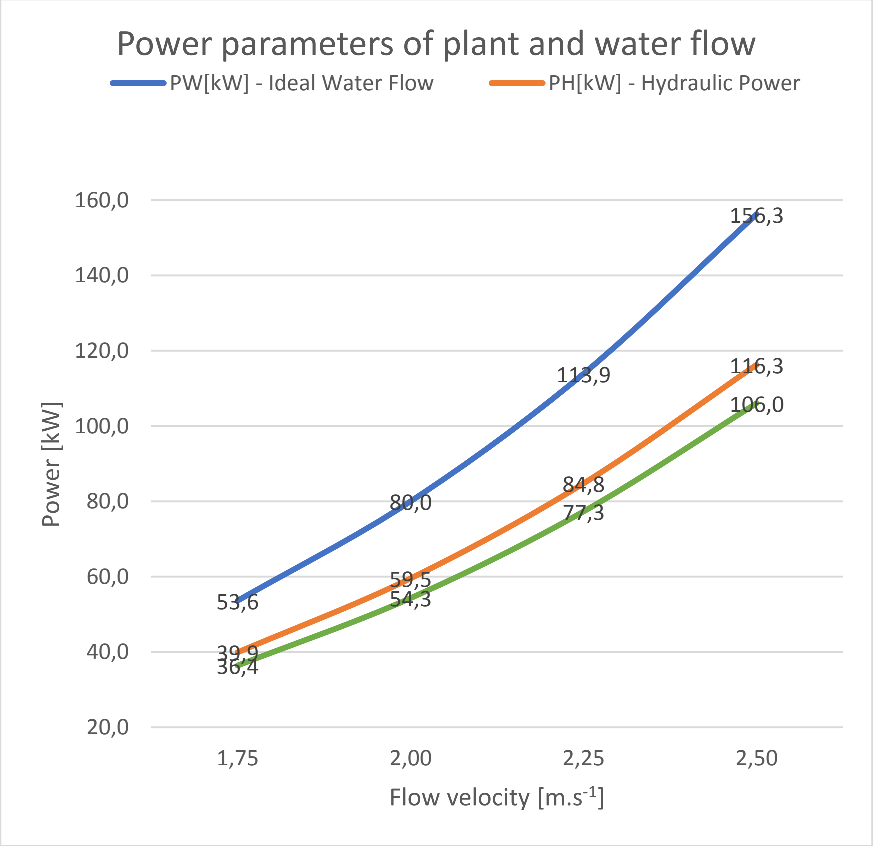


Figure 16
Dependence of hydraulic and electric power on water flow velocity

Based on the best results of the optimisation process, a conceptual design of a new hybrid power generation plant was developed using the already examined parameters and dimensions. To maximise the supplied electric power and the use of free space on the pontoon, it was proposed to supplement the equipment of the floating power plant as follows:

- Creating a light roof surface with an area of ~300 m^2 for placement of photovoltaic foils with a nominal output of 50 kW;
- Installation of two wind generators in a suitable place above the roof with a nominal power of 8 kW.

These sources of electricity will be sufficient to supply the operating and auxiliary systems of the pontoon with energy and thus will not burden the main generators. In the next steps of development, we propose to examine and implement the following measures for the construction of the hydroelectric power plant:

- Reduction of frictional resistance on the inner walls of the pontoon;
- Improving the water supply through the side inlets of pontoons;
- Further optimisation of the water wheel and the shape of its blades.

By implementing such measures, we can estimate that a 10% increase in the delivered electric power of the hydropower plant could be achieved. The planned model tests were not performed yet, but on the basis of the signed optimization, a similar functional power plant with a water wheel on the Mura River was built.

Conclusions

The results of the optimisation process confirm that it is possible to build up a floating small power plant using "classic" water wheels with higher efficiency. Compared to the exemplary single-tier type, which operates with 55% efficiency, the optimised type F should actually reach over 68% (after further optimisation steps). This significant increase was achieved with the following characteristics:

- Optimised pontoon shape, closed tunnel from below - even though the viscous resistance of the tunnel increases, the overall performance is improved because the water flow does not have the possibility to bypass the wheel from below;
- Optimised wheel geometry, shape, and arrangement of the blades - special shaping and division prevented possible transverse flow in front of the blades and managed to achieve a smoother run along with reducing the susceptibility to get in "oscillating states“ in CFD simulations when working in pairs;
- Two-tier wheel arrangement – the performed CFD analyses showed that with a single-tier arrangement it was no longer possible to significantly increase the efficiency, in the optimisation stage we chose the two-tier assembly. However, a simple doubling of the wheels would not bring the desired effect, due to the slowing down of the water flow in the tunnel with respect to the surrounding river flow. Therefore, it was necessary to specially shape the water wheels, the floating pontoon, and the tunnels;

- Supply of "fresh" water from the side of the pontoon to the rear wheel - calculations have clearly shown the advantage of side inlets on the pontoons. Where they were not used, the flow gradually slowed down and the simulation also showed a lack of water in the space between the two wheels (in extreme cases, oscillating states also occurred).

CFD analyses performed in the last phase were at the limit of the software's capabilities, were very demanding on the computing power of the hardware, and in some cases showed some instability. Therefore, all cases were analysed multiple times, if it was reasonable.

Acknowledgement

This research was funded by Institutional research of the Grant system of the Faculty of Operation and Economics of Transport and Communications, University of Zilina. This research is also the result of the Project VEGA No. 1/0128/20: Research on the Economic Efficiency of Variant Transport Modes in the Car Transport in the Slovak Republic with Emphasis on Sustainability and Environmental Impact, Faculty of Operation and Economics of Transport and Communications, University of Zilina, 2020-2022.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] Dvorak, Z., Rehak, D., David, A. and Cekerevac, Z: Qualitative Approach to Environmental Risk Assessment in Transport. Int. J. Environ. Res. Public Health 2020, 17, Art. no. 5494

[2] Hecht, J. S., Lacombe, G., Arias, M. E. and Dang, T. D: Hydropower dams of the Mekong river basin: a review of their hydrological impacts. J. Hydrol. 2019, 568, 285-300

[3] Xie, X., Jiang, X., Zhang, T. and Huang, Z: Regional water footprints assessment for hydroelectricity generation in China. Renew. Energy. 2019, 138, 316-325

[4] Mekonnen, M. M., Gerbens-Leenes, P. W. and Hoekstra, A. Y: The consumptive water footprint of electricity and heat: a global assessment. Environ. Sci. J. Integr. Environ. Res.: Water. Res. Technol. 2015, 1, 3, 285-297

[5] Kibler, K. M. and Tullos, D. D: Cumulative biophysical impact of small and large hydropower development in Nu River, China. Water Resources Research, 2013, 49, 6, 3104-3118

[6] Tiago, G. L., Silvia dos Santos, I. F. and Barros, R. M: Cost estimate of small hydroelectric power plants based on the aspect factor. Renewable & Sustainable Energy Reviews. 2017, 77, 229-238, DOI: 10.1016/j.rser.2017.03.134

[7] Euractiv: Assessment of Small Hydropower Plant. Available online: https://euractiv.sk/section/voda/opinion/male-vodne-elektrarne-su-horsie-ako-velke-dokazov-je-dost (accessed on 10.12.2020)

[8] Nogueira, M. F. M., Lima, C. U. S. and Ribeiro, R. R. P: The use of small hydroelectric power plants in the Amazon. Renewable Energy. 1993, 3, 8, 907-911, DOI: https://doi.org/10.1016/0960-1481(93)90049-M

[9] Bakken, T. H., Sundt, H., Ruud, A. and Harby, A. Development of Small Versus Large Hydropower in Norway– Comparison of Environmental Impacts. Energy Procedia, 2012, 20, 185-199

[10] Aguilar, S., Louw, K. and Neville, K: IHA World Congress Bulletin. International Institute for Sustainable Development (IISD) and International Hydropower Association (IHA) 2011

[11] Bueno, E. O., Mello, C. R. and Alves, G. J: Evaporation from Camargos hydropower plant reservoir: water footprint characterization. Revista Brasileira de Recursos Hídricos, 2016, 21, 3, 570-575

[12] Mekonnen, M. M. and Hoekstra, A. Y: The water footprint of electricity from hydropower. Value of Water Research Report Series. 2011, 51, 36

[13] American Rivers: Hydropower and Climate Change. Available online: https://www.americanrivers.org/threats-solutions/energy-development/hydropower-climate-change/ (accessed on 12.12.2020)

[14] Kadar, P: Small Hydro Plant Development. International Conference on Renewable Energies and Power Quality. Cordoba, Spain, 8-10 April 2014

[15] Illes, L., Kalina, T., Jurkovic, M. and Luptak, V: Distributed Propulsion Systems for Shallow Draft Vessels. J. Mar. Sci. Eng. 2020, 8, 667

[16] Douglas, J. F., Gasiore, J. M., Swaffield, J. A. and Jack, L. B: Fluid Mechanics, 5th ed.; Pearson: Harlow, UK, 2005

[17] Molnar, V: Computational Fluid Dynamics—Interdisciplinary Approach with CFD; University of Technology in Bratislava (STU): Bratislava, Slovakia, 2011

[18] Kudelas, D: Basic of Computer Flow Modelling and Visualization; Faculty of Mining, Ecology, Process Control and Geotechnologies: Kosice, Slovakia, 2017

[19] Mikušová, N., Stopka, O., Stopková, M. and Opettová, E. Use of simulation by modelling of conveyor belt contact forces. Open Engineering, 2020, 9, 1, 709-715, DOI: 10.1515/eng-2019-0070

[20] Castro, H., Putnik, G., Castro, A. and Fontana, R. D. B: Open Design initiatives: an evaluation of CAD Open Source Software. Procedia CIRP, 2019, 84, 1116-1119, DOI: https://doi.org/10.1016/j.procir.2019.08.001

Consumer Control Supportive Visualization

Gyenge Balázs [1], Szeghegyi Ágnes,[2] Szalay Gábor[1], Kozma Tímea[3]

[1] Department of Operation Management and Logistics, Supply Chain Management, Marketing and Tourism Institute, Faculty of Economy and Social Sciences, Szent István University, Páter Károly út 1, H-2100 Gödöllő, Hungary, gyenge.balazs@szie.hu, Gabor.Szalay@hu.bosch.com

[2] Keleti Károly Faculty of Business and Management, Óbuda University, Tavaszmező utca 15, H-1084 Budapest, Hungary, szeghegyi.agnes@kgk.uni-obuda.hu

[3] Institute of Management and Business Informatics, Budapest Business School, Buzogány u. 10-12, H-1149 Budapest, Hungary, kozma.timea@uni-bge.hu

Abstract: It is becoming increasingly clear to professionals and leaders, in the economic sphere, that logistics and logistical approach are now not only a service area, but also clearly, part of a competitive winning strategy. We can see the benefits of system, both in the supply chain and in the flow actions of value-creating processes. Lean, is one of the most comprehensive and respected philosophies of value-creating process development, which now weaves the organization of resupply through its wide range of tools, including external relations on more and more levels. We have good reason to ask, what can be used to make the internal and external development of needs even more transparent and plastic? It is extremely important, that the consistency of demand and production is established in a customer-centered approach, which must be experienced by all participants, even those who are not directly related to the customers or their needs. In the meantime, in the growing competition between companies, the product itself is no longer the most important, but the services associated with it, for which we will need additional production information with the highest efficiency and economies of scale. At the same time, customer expectations are changing at a faster pace than ever, which requires not only extreme flexibility and preparedness, but also immediate (up-to-date) information. Considering of all this, this study looks for the technology to help display and monitor consumption, in a controlled way. This helps targeting the goals and identify workers, within the current situation, thus demonstrating the significance of visualization, through live examples. Our main question is to look at the "whats" and "whys" of visualization and once knowing that, we will also be better equipped to perform the visualization.

Keywords: Visualization; Visual Management; Consumption Control; Logistics; Lean

1 Introduction, the Importance of Consumption Control in Today's Economic Market

In today's advanced, globalized manufacturing market, regions and national borders are no longer barriers to technology transfer. Large companies can access almost all technologies with sufficient capital and develop almost all skills among staff. Although there are significant differences in this area, the condition for staying in the market is noticeably pushed in the direction of how quickly and in what capacity we can serve customer needs and how cost-effectively we are able to do so. Consumer demand and continuous improvements in mass-customization, even in mass and series production, are shifting towards individual customized products, but quantitative fluctuations in customer demand are still very challenging. Although large customers (large suppliers, manufacturers, distributors) can reduce fluctuations with more accurate forecasting, to some extent through contracts, strategic cooperation and deliberate delay of orders, this problem can only be addressed effectively with a sufficiently advanced balanced or smoothing-pull system production scheduling.

Large companies using advanced production systems and producing customized products, and their senior colleagues, agree that push production, which accumulates large stocks, is no longer competitive [1]. To make finished and semi-finished products economically, assortment and continuously available, it is crucial to keep stocks at a low level, which will ensure the expected variance. To this end, a flexible production system, rapid changeover, fast lead times and predictable, stable technical (and human) capacity should be ensured on the technical side. From a non-technical point of view, it is necessary to clarify the forecasts by applying and/or developing the procedures and methods of the business, by defining inventory levels based on knowledge of the thorough process and customer behavior, and by providing and displaying accurate real-time information to the operating staff. One of the most effective ways to acquire knowledge to boost decision making or to develop a kind of decision support system, the visualization or so-called visual management method. In our study, we are looking for the technology to help display and monitor consumption in a controlled way and helps to target goals and identify workers with the current situation. We will try to demonstrate the significance of visualization through a live example. Our main question is to look at what and why to visualize it. And once we know that, we will also know how to do it. In the next section we'll continue to explore the importance of visualization in a specific industrial environment and look for opportunities for a visualization of consumer needs.

2 Literature Review on Visualization

Many companies took serious efforts and introduced different physical visual tools that have been implemented to facilitate performance measurement or other communication in different processes. There are a lot of different tools to make better complex knowledge and overview about any value creation processes, like visual process boards [2], visualize ERP outputs [3] or any tools to make a better control on multiple supplier inputs or material flows (see MRPs [4-7]). It is crucial to bring together the different viewpoints or using visual process control to boost the communication and cooperation among the individual participants and make faster and more effective run on work lots, resources and processes throughout their organization. These systems act as an extension to statistics, metrics [8], and in themselves may be considered as a dynamic measurement system as they provide instant feedback and can be used to predict a probable outcome. [3] "Visual process management tools have been mainly developed by LEAN practitioners as communication aids and are used to help drive operations and processes in real time" [3]. Visual management is a way to visually communicate expectations, performance, quantities, standards, or warnings in a way that requires little or no prior training to interpret. We also use visualization if we have a numerous data like in case of Big Data. Different characteristic of the data and usage makes different visualization types in analytical literature sources, e.g. large volume of data makes 'Volume type of visualization', in case of multi-format data presentation gives 'Variety' type, and high data processing speed makes 'Velocity' type necessary. [9] One of the current challenges confronted by organizations is how to improve the ineffective delivery of information to their workforces in close-range communication [10]. The expected result of visual management is improved operations at a work setting [3] [11] [12]. Main advantage of visual management is the ability to instantly show the current state, desired extent, tracking dynamic changes, distributions, detect problems, wastes, signaling or highlighting decision points to anyone that observes, within only seconds. Effective visual management uses unique visual signs to communicate in many ways and it requires no any additional explanation to understand. We may share, build in, warn, stop, prevent any information or abnormalities to improving control of guidelines, performance and quality. In analysis tasks, the analyst usually wants to access a whole data array, but the data or tables are not communicate well, cannot be interpreted deeply so we need visualization. The first level of visualization is reduction of data sets like classification and different statistical methods and modelling. These approaches operate with multi gradient data aggregation and filtration, based on the relatedness of objects in concrete dataset by one or more criteria. The second level of visualization to use any 2D or 3D diagrams which makes the ability to see trends, similarities, seasonality, fluctuations, or margins for interpretation and so on. (e.g. temporal, hierarchical, network, multidimensional, geospatial diagrams) The effective visualization does not stop on only static image or data visualization, so the above problems become more significant in dynamic visualization. I believe that real-time tracking or dynamic visualization, can be the third level of visualization when the time has a special aspect to interpreting the meaning of data.

Graphical thinking is a very simple and natural type of data processing for a human being, so, it can be said, that image data representation is an effective method, which allows for easing data understanding and provides enough support for decision making [9]. Graphical data visualization increases the level of perception because human being loses the ability to acquire any useful information in any overloaded situation.

The collated experiences and implementation themes that have made as set of guidelines are crucial for new implementations and possible novel innovations [9] which boosts the communication and better control in a LEAN manufacturing process. Moreover, visual management can effectively be implemented in many other areas not just manufacturing [3] [13] (see examples on: Rolls Royce, Airbus, or other areas e.g. construction organizations, IT and software, service, commercial, educational, healthcare, or governmental service) [14-17]. In LEAN environment the visualization shorten lead times, reduce inventory, develop a safe working environment and even boost your profits and income. In the following study step-by-step we will elaborate a complete concept of gauge visualization of Kanban needs and guide for visual decision making.

3 Material and Method, LEAN Summary and Principles of Consumption Control

3.1 The LEAN Methodology

First of all, we would like to stress that the LEAN methodology is not "just" a production technology system for us, in which certain methods of procedure are defined or required. Like Japanese professionals, we stress that, above all, it is a more comprehensive philosophical concept that always favors solutions that provide the greatest usefulness and conformity at the moment (regardless of its origin or authors), so openness and willingness to give up previous innervation, traditional patterns, are important for its application and full understanding and making room for change.

By applying Lean's principles, you can achieve efficient and economical operation. The Lean methodology focuses on the process and not on the output of the process [18] or nor the function. At the same time, the starting principle is that all elements of production should be re-thought and developed on the basis of customer needs.

In our methodology used below, we look at the processes of a multinational company as case-by-case examples, and then we carried out a secondary literature analysis, comparing the process and the area point by point with theoretical concepts, adapting them to local specificities with critical observations. First, let us review the interpretation of the principles based on the literature. According to one of the first newsreel and interpreters of the Lean concept (Womack – Jones, 1996) [19], the principles of Lean philosophy are the following, which we have paired with the most common tools according to Liker (2008) [20]:

- The concept of value and the definition of value – in which they stress that only activities and product elements are valuable, that are valuable to the buyer and what the buyer is willing to pay for, which also reveals real losses. To analyze it: analyze customer needs and how it is consumed.
- Value stream (value process) – when mapping the entire activity, it is clear which activities and time needs create value in the sense mentioned above and how they form a structure and chain. To analyze: you need to analyze the sources of loss, often used in the five why technique and value stream analysis (VSM).
- Flow – the more we talk about a process system, the more important is that the product and material flow is balanced, i.e. free of fluctuations, since bottlenecks and congestions hinder fast cycle times and high permeability at the level of the entire system. For analysis and development, in the spirit of JIT, pace-time planning, rapid changeover (SMED), built-in quality (JIDOKA), balanced production (HEIJUNKA), and developing one-piece flow are used.
- Pull – the key to reduce stocks dramatically is that every flow must be linked to demand and not the other way around, so there is a flow only if there is an order demand or signal in the balanced production chain, meaning there is no production for stocks except for intermediate buffers. For analysis and development, developing pull production (PULL), ensuring fast and automated supply is created by using KANBAN technology, and value flow analysis (VSM) is also often used.
- Improvement – continuous development and introduction of Kaizen, ensures small-step development in all areas, at all levels of the company, effectively, quickly and motivated, including the zero error principle. For analysis and development, we often use process standardization, continuous improvement KAISEN and transparent 5S technology.

As you can see, the Lean approach consists of stacked philosophies, principles and less specific methods or a wide variety of tools.

"If you learn only methods, you will be tied to your methods, but if you learn principles, you can devise your own methods." – Ralph Waldo Emerson

Toyota's production system (TPS) is the source of LEAN principles and is now one of its fore-holdings. If, although primarily used in the production area (now a very multi-layered methodology, which has now become part of standardization, idea channeling, motivation, and atomization, together with a number of other commonly used methods, which, of course, can be used in other areas, such as the production of services. We must also bear in mind that Toyota itself designates humans as a fundamental value and the involvement of workers, by its interpretation we often see very big problems in practice and in our different corporate cultures.

Professor Blanchard (2010) [21] clearly highlights all this and not only the technological foundations (radically new perspectives) of the Lean system:

- All employees must be empowered to develop their company according to their ability and responsibility.
- Toyota's production system is based on a philosophy of continuous development and respect for people.
- Lean management is a waste eliminating strategy, not a cost-cutting strategy.
- Lean's practices must be closely linked to the process of the company's supply chain.

High-impact and all-in-one thinking on continuously implemented development – Kaisen which is the built-in quality supplemented by the JIDOKA and JIT (just in time) principles, all of which aim to significantly reduce or eliminate losses, mainly through modifications that can increase lead times and the efficiency, flexibility and responsiveness of the system. [22] [23]

In their study, Pónusz and Sáska (2015) [24] examined how the creation of basic standards and the development of a well-transparent (visualized) working environment lead to the effective implementation of Lean philosophy. In the following illustration, Toyota's own concept was further developed in a house-like diagram, grouping Lean's philosophies, principles, methods and tools from strategy to operations, where it is very important that even "principles and philosophies" are grouped around methods and tools, the practical implementation, which ensures the desired greater efficiency. Based on the previous model of Liker (2008) [20].

The Toyota house symbol was first illustrated in the form of a house by Fujio Cho, a student of the famous engineer, Taiichhi Ohno. The purpose of the depiction was to show that, like the house, it is a structural system that needs strong foundations and supporting walls so that the goal, i.e. the roof, can be created. By analogy, a single weak link can weaken the entire system. Lean is a process management approach whose main goal is to create customer value and eliminate waste. “Lean provides the customer with fewer employees, fewer devices, less time, and less space, with fewer resources (more) value.” [25]

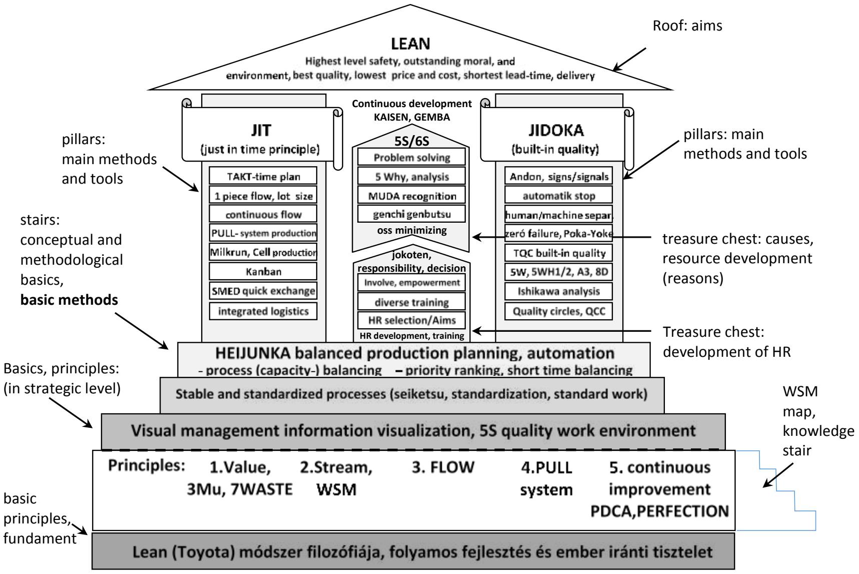


Figure 1

Lean tools and principles built on each other, as „Lean-house” concept (source: based on Liker (2008) [20] but developed and reconstructed figure, and recommendation)

It is worth noting, that there may be many types of waste, since value-creating processes that do not fulfill some expectations, processes that do not add value, but are also necessary (MUDA I) and, processes to be minimized that do not add value but are necessary and the processes for eliminating losses that do not add value and are not necessary (MUDA II) In order to recognize all this, the necessary mentality and attitude are not easy to achieve, which have a re-effect on philosophical foundations. In some organizations, not only technology is evolving, but also the attitude, which makes it possible to recognize certain types of losses and accentuates at a particular point in development. Based on Taiichi Ohno and Womack (1996) [19], we can create the following categories according to our own interpretation, supplemented and significantly rephrased:

7 wastes of Lean – 7 main sources of loss (in extended generalized interpretation):

1) In relation to over-stocks (inventory): according to our interpretation, all inventory costs are originally losses, since they do not generate value while being in stock, but unreasonably more stocks than their smaller quantities are certainly a competitive disadvantage, i.e. a loss. Determining the extent and amount of it is very challenging which cannot be simplified into a mathematical problem (in fact, the challenge for the human intellect).

2) Transportation loss: in the case of transport, it is not always the case that the buyer is willing to pay for it, thus significantly increasing the lead time of the value-creating processes and can also implement a number of unnecessary surpluses (e.g. extra road, extra time).
3) Defects and waste: i.e. all forms of defects are losses, but fewer people take into account their consequences, such as other production distractions, which are additional sources of loss, not to mention a disgruntled customer and a decrease in loyalty.
4) Waiting time: any time spent in a production system that does not generate value while resources are running out or, although not running out, capacity utilization does not reach a certain level and therefore production takes more time than necessary.
5) Over-processing: Any unnecessary activities or any work performed for which the customer no longer pays, but still necessary, e.g.: no matter how many times a product has to be checked, touched or repaired if the price is the same.
6) Motion: every wasteful movement any movement that someone has to make, but is not necessary, or in certain cases the whole movement may be completely unnecessary and it is only and exclusively subject of the procedure or arrangement.
7) Over-production (or over-ordering): any surplus to which there is no registered external or internal need, all that, what is not compatible with the pull principle.

In our investigation, our basic problem is the extent method of adjusting to demand levels. There are many approaches in the literature, from mathematical optimization to dynamic programming or simulation modelling. According to one possible concept, it is worth getting to know the distribution of needs as accurately as possible and we should try to adapt to them with the most accurate planning. The other possible concept is to increase the potential performance and flexibility of the service to a maximum and strive for an "immediate" response. The former is called MRP-based supply planning, while the latter is approached in JIT and LEAN systems. For the latter systems, the optimal solution is to keep inventory levels low and fast cycles, although it may be necessary to maintain and plan certain safety stock levels in both basic systems. In our case, we base our production on the concept that we try to adapt production as best we can to customer needs on a kind of synchronous production basis. The more we move towards flexible and synchronous production, the more problems arise in the non-equal utilization of capacities, which generates loss of efficiency or additional costs, or may result in periods when demand exceeds available capacities at any given moment, causing “lost” income (i.e. alternative costs). In the following illustration, we can see that in the case of large fluctuations, capacity scaling can be problematic for a number of reasons.

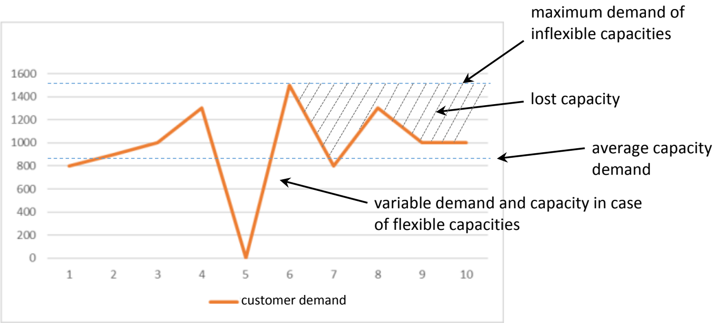


Figure 2

Volatility of demand (source: own editing from 2018 data of the examined company)

Some capacities cannot typically be changed flexibly (they should be scaled to maximum in a service strategy), while others may be more flexible and more free to follow changes in demand (typically with an extra premium cost, which is insured in relation to the constant level of capacity). Keeping inventory levels low and adapting production flexibly to customer needs does not always allow us to use capacity, which means additional costs for machinery and human resources, but if we keep capacities below the principled maximum, it is also possible that we will not be able to meet the demand peaks or even transfer them to a later date.

A smooth production schedule will make it possible to distribute needs more sensitively over time and thus make fuller use of capacities and reduce inventories further, at the cost of more transitions, more transition time and transition costs, provided that it can be reduced to a small level. Thinking in reverse, we can also plan the level of flexible capacities, which can be called a smoothed capacity. This way, you can manage both low inventory levels and fuller capacity utilization. After clarifying the forecasts and a more thorough assessment, it was established what distribution unmatched customer needs might appear in a possible next period. (These results are not referred to in this paper.)

If the production value chain is sufficiently complex, the completion cycles (beat, drum) of the bottleneck process will usually play the pace and “pulsation” of the entire system. In practice, we smooth out the needs for a period determined by forecasts and planned customer “takeaways” and provide this as a production plan, i.e. a requested capacity plan. Plans must be broken down to item number level.

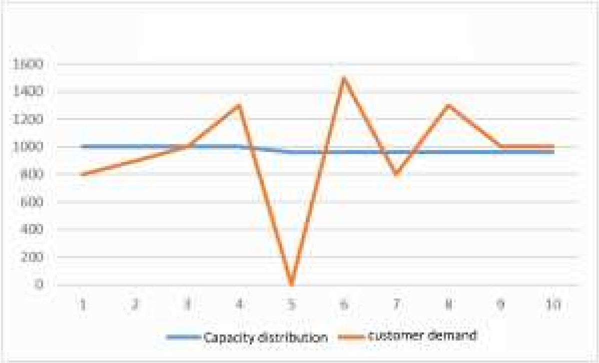


Figure 3
Smoothed capacity (source: own editing from 2018 data of the examined company)

The principle of consumption-based capacity control (i.e. consumption control) is broken down by item number like:

1) Process flow (VSM)

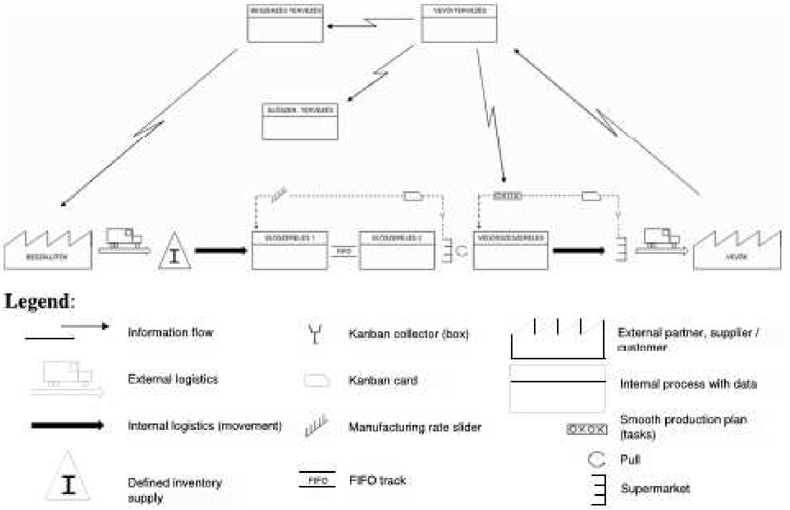

FIFO
Legend:
Information flow
Kanban collector (box)
External partner, supplier / customer
External logistics
Kanban card
Internal process with data
Internal logistics (movement)
Manufacturing rate slider
OXOX
Smooth production plan (tasks)
I
Defined inventory supply
FIFO
FIFO track
Pull
Supermarket

Figure 4
Smoothed capacity (source: own editing from 2018 data of the examined company)

2) Interfaces:

Production is separated from suppliers with "defined" stocks and from customers by Supermarket. The delivery cycles differ between our supplier and our customer. Two major areas are distinguished in internal processes (smaller processes within areas are less relevant and therefore not included):

- Pre-assembly: pre-assembly1 and pre-assembly2 due to the complexity of the process, which essentially carry out the same process, but with different steps
- Final assembly

The following interfaces can be found in the production process:

- Raw material warehouse → Pre-assembly 1 and 2: the raw materials are seconded through an automatic system, according to your production plan according to your needs. The supply of materials is carried out by automated equipment on a specific cycle.
- Pre-assembly 1 → Pre-assembly 2: the two pre-assembly lines are connected by a FIFO track (slider).
- Pre-assembly 2 →Final assembly: supermarket kit is located between the two segments, this provides the possibility of separation. The existence of the Supermarket is also explained by the production of several different finished products from the pre-assembly item numbers and by the work of several different end-assembly lines from a pre-installed item number.
- Final assembly pushes the finished product to the finished product Supermarket.

3) Flow of information and materials

The flow of information can be divided into two parts:

- At specific intervals, system-wide data for calculations and planning within the design horizon.
- Information flow related to material flow and inventory levels (via Kanban circles).

System data shall be provided at fixed intervals, which may be weekly, bi-weekly, monthly, bimonthly, etc. The longer this design time horizon is, the more secure the customer forecasts, the more predictable the production system and the more predictable customer behavior.

In the value stream outlined above, the flow of information is:

- Customer planning receives orders from customers at item number level.
- Customer planning provides information about:

  1) In the direction of purchase → handling of raw material orders.

  2) Towards final assembly → production plan smoothed on the basis of capacity data and customer takeaways.

3) In direction of pre-assembly planning → final assembly smooth production plan + ABC analysis pre-assembly broken down into item numbers.

- Pre-assembly planning:
  - Pre-assembly capacity planning
  - ABC analysis: separation and distribution of main running and exotic part-products on the basis of capacities.
  - Kanban calculation and definition of supermarket stock levels (minimum maximum) for main running products, by part numbers.

Input information refers the value stream at several points and induces manufacturing activity. However, in addition to information from outside, there is also internal information, some of which are linked to physical material flow:

- The Kanban from supermarkets, until it is re-manufactured, is present as information in the system and flows backwards.
- At the stage of the process at which the product starts to be manufactured, Kanban information is stored on the products or their containers and the information and the product as a material go together through all processes all the way to the Supermarket until the customer or consumer "pulls" the product away from the Supermarket.

3.2 Data and Information Needed for Operation and Analysis of Visualization

One of the most important criteria to be able to operate consumption control is to be aware of the inventory and demand data that affects production in the short term, as well as sufficient information to make our decision more rational in some way. Experience has shown that the following 6 information are essential to create a real-time picture of operation:

1) Current stocks that are also available to the customer or consumer: item number and quantity, and location.
2) Not yet available to the customer, but stocks in progress, in production, and their degree of completeness. This is so-called: "pre-scheduled, moving stock."
3) Information on stocks already used by customers but not yet started in progress (e.g. re-rotated Kanbans awaiting production). If we are aware of the needs and capacities of our customers (and in an industrial environment, the least what we can do is to have a strong forecast of the expected needs, which is likely to be in sync with the stocks of the customers).
4) Pre-planned customer service or "takeaway".

5) Planned downtimes in production (maintenance, breaks, etc.).
6) Current capacity utilization indicators.

Lean management has a set of tools designed to visualize deviations from standards, expected ranges, goals. A visual display is more expressive than any data or text. The portfolio of visual management is full of marking and signaling techniques that can be used. Such visualization systems include the Kanban, the production island Andon and standardized work, for example, as they make the thing, we need to pay attention to immediately visible and perceptible. These tools not only consciously cross lean systems, but they also provide information to those involved in the process and managers immediately when used, if they experience any differences.

Visuality is a basic concept (as shown in Figure 1) that allows the losses in processes to be visible and helps the decision-makers concerned determine the most appropriate form of intervention. It makes it easy for anyone to decide whether processes are going well or not, and then if not, what pre-drafted scheme is needed to intervene. An essential part of the day’s work, it provides information about the current state of the production process (progress, delay, material shortage, machine failures, inventory status, etc.). It is also an effective tool for improving the production system by helping to identify problems and identify the location of problems. The first step in applying visual control is to create some sort of order that ensures transparency, so it is often created after 5S. An important criterion for Lean's production management tools is to comply with the principles of visual control, i.e. to provide easily understandable, clearly visible information on the state of production. [26]

The question arises as to what, where and how to display it in our case. Also, an additional question is who needs the impression, who will be the potential users?

All these issues are interconnected and functionally define each other. For example, you might want to display data at the point of production where things happen or where intervention is relevant. A common solution is to display at the very beginning or at the very end of the process if a process is short and well-bounded. If digital solutions are available, it is a good idea to display them digitally because they can appear at multiple points in production (“anywhere”). The question “how” means not only the digital term, but also the most expressive representation currently known, which has the greatest recognizable or expressive power for users. If this information is already collected in a bunch, it will be the colleagues working in the process and the direct managers responsible for the activity who need to know them, as well as those responsible for planning the activity, in close cooperation with the previous two. With all this in mind, there is only one question left: What should we display?

4 Results and Discussion

From a view point of view, the data can be divided into two parts:

1) Real-time data: i.e. whenever I look at the board/monitor, I immediately have a picture of how the production is currently going. This is necessary to recognize whether everything is going well, or if there is any disruption that I need to respond to in a crisis. If the necessary data are all displayed at the same time and can be easily interpreted in their context, then it is possible to react quickly and make relevant decisions.
2) Historical data (statistics): all that has happened in the past, which can be natural or cumulative values. These impressions are important for improving the system and following the goals. We can use the statistics to make analyses and make estimates for the future.

With this theoretical approach in mind, I would like to outline the revised concept of the specialized consumption for implantation (pre-assembly) SMT control visualization of an automotive supplier company and to provide framework to the analytical procedures.

When designing a visual, the criterion is that the graphic/image/signal displayed is sufficiently expansive: it should be immediately able to provide a comprehensive knowledge about whether the current situation is appropriate or inappropriate. If you want a good visualization, you should also strive to make it easy to understand especially for users and even for those who do not have a deep knowledge of the process or production line (smart figure concept). For example, you can display charts instead of numbers, even if specific numbers remain "data-minable" (drillable). In the same way, we want to avoid anything that needs to be explained in text, because the text is subjective to the figure is clear. The time dimension of the representation is also important, in which the on-time (current) state becomes very important, especially when the management aspect comes to the fore. Also important are the colors, which are very expressive (red – not good, green – good, yellow – warning, etc.). Diagrams and symbols must wear their purpose, which may be indication, separation, coding, explanatory and interpretative meaning. It is a good idea to design different display levels and gradients, depending on the purpose and user of the report.

With all this in mind, the following visualization has been introduced. The first-round display is as follows:

1) On-time inventory tracking, according to data extracted from a database in every 30 minutes.
2) Follow supermarket entry and pick-up, broken down by day.
3) Follow supermarket levels, broken down to varying divisions, over the period of 30-minute data.

Second-round display:

1) Display of current inventory level data
2) Supermarket entry and pick-up tracking, exact entry and take-away status.
3) Follow supermarket inventory according to the levels
4) Production plans and monitoring (historical data not described in this study)

4.1 On-Time Inventory Tracking

During Kanban circle calculation, for example, it is important to have a minimum and maximum level of inventory that activates specific reactions or escalations. For example, in the event of a decrease, the minimum inventory level triggers automated replenishment and escalation. In this case, the maximum level limits the total quantity of the item number that can be manufactured, and if there is no customer or user “pull” (order), a maximum of this amount of inventory can be created (upper buffer limit). These exact boundaries based on preliminary calculations provide the framework for displaying the current inventory level in each storage location. We want to mark the different levels and the current quantity in a distinct color, which has already been introduced in the production area under investigation. For the first-round display, it is sufficient to display only as a chart, while the second-round display must show the exact data: what is the minimum, what is the maximum, and what is the quantity of the current inventory level.

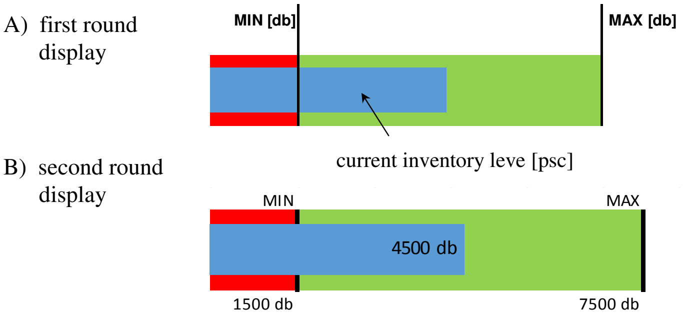


Figure 5

Inventory tracking visualization (source: own editing of the examined production area)

Data sources:

- The minimum and maximum levels are recalculated periodically from the data of the previous Kanban circle
- The current inventory level, can be retrieved from the always valid management information system used in the production area – samples

should preferably be taken as quickly as possible, but at least more frequent than customer takeaways and/or synchronized with it in every 30 minutes.

Users:

- When assessing the current inventory level, the production planner
- When assessing the current inventory levels and starting a reaction/escalation, the direct management level, such as a shift manager.

4.2 Supermarket Entry and Pick-Up Tracking

In order to fit the shape of production and demand better, we need to look at the in and out flow of the Supermarket. We need to know what kind of load and picking have taken place during a specified period, which can be illustrated on the same scale and daily breakdown as a specially designed inflow and outflow chart. It is a good idea to show two or three weeks in the chart to create a bigger picture. In our experience, it is also very graphic if the figure shows the planned average pull.

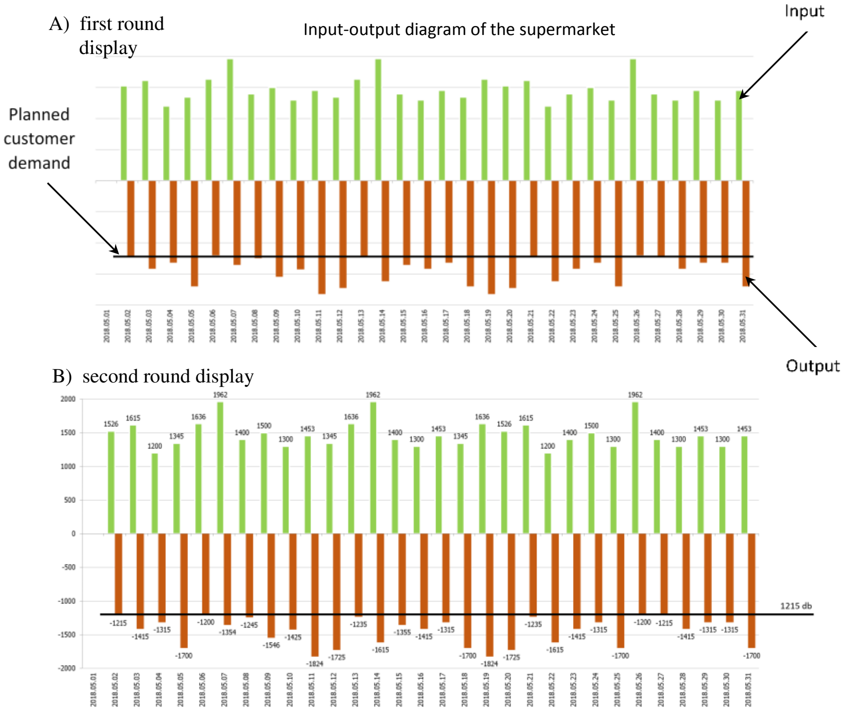


Figure 6

Supermarket loading and unloading visualization (source: own editing of the examined production area)

For deeper analysis, you may need to display the data in more detail, even in intra-day distribution. This is necessary where intra-day distribution is relevant.

Data sources:

- The planned customer (or consumer) takeaway can be derived from the smooth production plan of the final installation;
- The data displayed in daily (or intra-day) breakdowns can be extracted from the integrated management information (ERP, TPS, EIS, CAPP, ect.) system used in the production area.

Users:

- In the case of customer takeaway analysis, the production designer;
- If the minimum or maximum level is reached when the quantity and time of the customer's takeaway is reviewed and broken down the direct management level, such as the shift manager;
- Higher management level for customer takeaway, for analysis of the production volume and for system-level intervention.

After the visualization of the Supermarket loads and pic-ups, we consider the process-tape-like display to be a classic but very useful representation, illustrating the evolution within the desirable range in a clear form, as well as meeting our historical data needs about how the inventory has changed. For this visual, it is a good idea to use the same minimum and maximum levels that you use for the real-time inventory visual. For the sake of reality, it is advisable to tune the density of the data recording to the frequency of the takeaway, which in our case will be 30 minutes. In the final visualization, the process tape is displayed as a chart that is spread out so that the deviation can be better observed in one direction or another within the desired range.

A) first round display

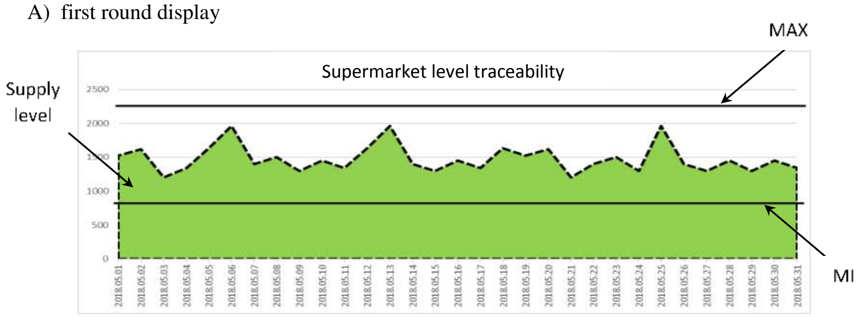


Figure 7

Supermarket level traceability visualization (source: own editing of the examined production area)

For data sources, the real-time inventory level tracing data sources are the same. In the case of users, the production manager and management levels are relevant for us, and in the case of the latter for the assessment of systemic interventions.

Conclusions and Suggestions

During the work herein, we came to the following conclusions concerning the basics of a good and desirable, visualization:

- It is called a good system and a good visualization that ensures transparency, which is also the basis for additional key criteria, which have been formulated as a kind of directive.
- A good visualization must be such that, the decision-making person can see the differences immediately, be able to draw conclusions, depending on the aspect or perspective from which they are looking, i.e. who the person in charge is and what their task is.
- A good visualization must be different for different stakeholders, which can mean scalability (gradient) or different levels in practice, allowing different perspectives.
- A good visualization is the basis for quick and effective decisions, which should not be complicated or too complex, or probable and enabling multi-decision, it should always be clear.
- A good visualization must be able to test causal relations or, in the case of further questions, to acquire a further deeper level of knowledge (data mining, drill).

On this basis, we have managed to create a visual for the area in question that meets the examined theoretical background, the specifics of our organization and the criteria listed above and reviewed in the analysis. In our proposal, we stress, and in agreement with management, we represent that real-time (gauge type) visualization and backlog on the right scale should be prioritized in a visual that facilitates everyday activity and operation in order to facilitate operational decisions by displaying it in a sufficiently simple way (the simpler, the better).

In this work, we have described an entire development process, including some theoretical issues and their solutions. This application sample contributes to development of other companies with similar problems and could directly generate productivity improvement for many companies. These concepts can be used by companies of any size, including small and medium-sized enterprises. There is a wide range of manufacturing efficiency improvement methods available to the companies, such as Just in Time (JIT), or a range of lean manufacturing tools, but this method doesn't need notable investment or any reorganizing. The limitation of our methodology is a continuous and not so fluctuated supply, which is common in a LEAN environment, but perhaps, not so typical in other industries. We recommend this kind of visualization in all cases where the constant control and intervention is required.

The effective use of LEAN principles and applicability, incorporation or acceptance, within operations, is a major problem for many companies. This is why it is so important to conduct scientific research on the best practices. Therefore, we have developed a strict methodology, that consists of clearly defined steps, based on an analysis of Kanban needs. This solution identifies the key visualization factors for supply and provides the basis for a detailed examination of the related production efficiencies.

References

[1] C. Dawson and A. Henley, "Push versus pull entrepreneurship. An ambiguous distinction?" International Journal of Entrepreneurial Behavior & Research, Vol. 18 No. 6, pp. 697-719, 2012, https://doi.org/10.1108/13552551211268139, downloaded: 23.09.2019

[2] U. Bititci, P. Cocca and A. Ates, "Impact of visual performance management systems on the performance management practices of organisations", International Journal of Production Research, Vol. 54, Issue 6, pp. 1571-1593, 2016, DOI: 10.1080/00207543.2015.1005770

[3] G. C. Parry and C. E. Turner, "Application of lean visual process management tools", Production Planning & Control, Vol. 17, Issue 1, pp. 77-86, 2006, DOI: 10.1080/09537280500414991 (first published online: 21 Feb 2007)

[4] P. Burcher, "Closed- Loop MRP", Operations Management, Vol. 10, and Strategic Management Journal, 22 January 2015A, https://doi.org/10.1002/9781118785317.weom100188

[5] P. Burcher, "Netting Process in MRP", Operations Management, Vol. 10, and Strategic Management Journal, 22 January 2015B, https://doi.org/10.1002/9781118785317.weom100188

[6] P. Burcher, "Safety Stocks in MRP", Operations Management, Vol. 10 and Strategic Management Journal, 22 January 2015C, https://doi.org/10.1002/9781118785317.weom100195

[7] A. Harrison and M. Lewis, "JIT and MRP/ERP", Operations Management, Vol. 10, 2015, https://doi.org/10.1002/9781118785317.weom100028

[8] Y. Eaidgah, A. A. Maki, K. Kurczewski, and A. Abdekhodaee, "Visual management, performance management and continuous improvement: A lean manufacturing approach", International Journal of Lean Six Sigma, Vol. 7, Issue 2, 2016, DOI: 10.1108/IJLSS-09-2014-0028 (first published: 6 June 2016)

[9] E. Y. Gorodov, and V. V. Gubarev, "Analytical Review of Data Visualization Methods in Application to Big Data", Journal of Electrical

and Computer Engineering, Vol. 2013, Article ID 969458, p. 7, 2013, https://doi.org/10.1155/2013/969458

[10] N. Bilalis, G. Scroubelos, A. Antoniadis, D. Emiris, and D. Koulouriotis, "Visual factory: basic principles and the 'zoning' approach", International Journal of Production Research, Vol. 40, No. 15, pp. 3575-3588, 2002 (first published online: 14 Nov 2010) https://doi.org/10.1080/00207540210140031

[11] C. Herron and P. M Braiden, "A methodology for developing sustainable quantifiable productivity improvement in manufacturing companies", International Journal of Production Economics, Vol. 104, No. 1, pp. 143-153, 2006, https://doi.org/10.1016/j.ijpe.2005.10.004

[12] Bhasin, S., "Lean and performance measurement", Journal of Manufacturing Technology Management, Vol. 19, No. 5, pp. 670-684, 2008, https://doi.org/10.1108/17410380810877311

[13] A. Tezel, L. Koskela and P. Tzortzopoulos, "Visual management in production management: a literature synthesis", Journal of Manufacturing Technology Management, Vol. 27, No. 6, pp. 766-799, 2016, https://doi.org/10.1108/JMTM-08-2015-0071

[14] S. Liff and P. A. Posey,"Seeing is Believing: How the New Art of Visual Management Can Boost Performance Throughout Your Organization", Publisher: AMACOM, New York, NY. 2004 (first published on amazon: February 27, 2007) ISBN-13 : 978-0814400357, DOI: 10.5860/choice.42-4137

[15] T. Joosten, I. Bongers and R. Janssen, "Application of lean thinking to health care: issues and observations", International Journal for Quality in Health Care, Vol. 21, No. 5, pp. 341-347, 2009

[16] Z. Radnor, "Transferring lean into government", Journal of Manufacturing Technology and Management, Vol. 21, No. 3, pp. 411-428, 2010

[17] M. O. Ahmad, J. Markkula and M. Oivo, "Kanban in software development: a systematic literature review", 39th IEEE Conference on Software Engineering and Advanced Applications (SEAA), pp. 9-16, 2013

[18] K. J. Caterall, "A Lean view on an Eastern Cape Logistics Service Provider" Publisher: Nelson Mandela Metropolitan University – Faculty of Business and Economic Sciences, https://vital.seals.ac.za/vital/access/manager/PdfViewer/vital: 8713/SOURCEPDF?viewPdfInternal=1, p. 129, 2008

[19] J. P. Womack and D. T. Jones, "Lean Thinking Banish Waste and Create Wealth in Your Corporation" Publisher: Simon & Shuster, New York, p. 400, 1996, ISBN: 9781439135952

[20] J. K. Liker, ”A Toyota-módszer: 14 vállalatirányítási alapelv, (The Toyota Way – 14 management principles),” (in Hungarian edition by T. Z. Polyánszky) Publisher: HVG, Budapest 2008, ISBN 978-963-9686-43-4

[21] D. Blanchard, “Supply Chain Management – Best practices”, Publisher: Hoboken, New Jersey, John Wiley & Sons, Inc. p. 215, 2010 (Second edition)

[22] B. Gyenge, H. Szilágyi and T. Kozma, “Lean menedzsment alkalmazása szolgáltatóvállalat esetében, (Lean management in case of logistic service provider company),” Vezetéstudomány (Budapest Management Review), No. 4, p. 46, Budapest, 2015, http://gazdalkodastudomany.uni-corvinus.hu/index.php?id =59089

[23] T. Kozma and M. Pónusz, “Ellátásilánc-menedzsment elmélete és gyakorlata – alapok, (Supply Chain Management principles and practice),” Publisher: Károly Róbert Kutató – Oktató Közhasznú Nonprofit Kft., Gödöllő, 2016

[24] M. Pónusz and Zs. Sáska, “Lean menedzsment eszközök gyakorlati alkalmazása Folyamatmenedzsment kihívásai, (Practical application of Lean management tools Challenges of process management),” in Döntési pontok, kapcsolatok és együttműködési stratégiák a gyakorlatban (Decision points, relationships and collaboration strategies in practice). Publisher: Szolnoki Főiskola (Szolnok College later: Naeumann János University), Szolnok 2015, ISBN 978-615-5570-02-5, http://fulltext.szie.hu/jadox/portal/displayImage.psml?docID=15876&secID=17291

[25] K. Demeter, “Termelés, szolgáltatás, logisztika. Az értékteremtés folyamatai, (Production, service, logistics. Value creation processes,)” Published: Wolters Kluwer Kiadó, Budapest, p. 159, 2014

[26] J. Kosztolányi and G. Schwahofer, “Lean szótár, (Lean dictionary)” Publisher: Kaizen PRO Kft. Budapest, 2012

An Analytical Solution of a Multi-Winding Coil Problem with a Magnetic Core in Spherical Coordinates

Hüseyin Yıldız[1,4*], Erol Uzal[2], Hüseyin Çalık[3]

[1]Istanbul University – Cerrahpasa, Department of Mechanical and Metal Technologies, Buyukcekmece, 34500 Istanbul, Turkey, huseyin.yildiz@istanbul.edu.tr

[2]Istanbul University – Cerrahpasa, Department of Mechanical Engineering, Avcilar, 34320 Istanbul, Turkey, euzal@istanbul.edu.tr

[3]Giresun University, Department of Electrical and Electronics Engineering, 28000 Giresun, Turkey, huseyin.calik@giresun.edu.tr

[4]Girift Technology Co., Avcilar, 34320 Istanbul, Turkey

*Abstract: Nowadays, technology is advancing rapidly in parallel with developments. The traditional concept of machine loses its function and is replaced by particular devices with spherical geometry such as spherical electric motors and brain stimulation systems. Consequently; the calculation of self-inductance (L_{ii}) and mutual inductance (M_{ij}) coefficients in the spherical coordinate system with analytical or semi-analytical methods has become one of the major research topics in recent years. In this study, the terms of **B**, **E,** and **A** for multi-winding coils were calculated analytically by using the single-winding coil approach. The same geometries were calculated with the assumption of axial symmetry in ANSYS Maxwell using finite element analysis (FEA). The obtained results with the FEA and analytical calculations were compared. Finally, two concentric coil geometries with magnetic core with radius r_1 were determined, the variation of the self-inductance (L_{ii}) and mutual inductance (M_{ij}) coefficients of the determined geometries based on the γ angle in spherical coordinates were calculated analytically. Simulation studies were conducted by creating 3D models of coil geometries in ANSYS Maxwell program. An experimental setup that can be produced with 3-dimensional (3D) printers has been designed and constructed compatibly with the determined geometries. The variation of L_{ii} and M_{ij} coefficients with γ was studied experimentally after the production of necessary coil windings. At the end of the study, it was observed that the analytical results collected for the mutual inductance M_{ij} and the self-inductance coefficient L_{ii} were consistent with each other by comparing the FEA and experimental results.*

Keywords: Spherical coordinates; Magnetic field; Maxwell's equations; Analytical solution; Finite element analysis (FEA); Mutual inductance; Self-inductance

1 Introduction

Recently, in parallel with technological developments, the traditional concept of machine has lost its function and replaced with very particular devices. By the advance of accelerator technology, the development of brain stimulation systems, the increase in wireless energy transfer studies, and the development of spherical electric motors, analytical calculation of self-inductance (L_{ii}) and mutual inductance coefficients (M_{ij}) in different coordinate systems has become one of the major research topics. Therefore, the calculation of the inductance coefficient by analytical or semi-analytical methods has attracted great attention in recent years. Much progress has been made in the solution methods in Cartesian and cylindrical coordinates [1-2]. The majority of the studies conducted have an entirely circular geometry. Calculations for circular loops [2-4], cylindrical discs [5], and cylindrical shells [6] can be found in the literature [1]-[7]. Some studies also examine the analytical form of the mutual inductance coefficient for non-coaxial thin and thick coils [8]. These suggested solutions are usually expressed as elliptic integrals, Legendre functions, or Bessel and Struve functions.

In the design of conventional machines, analyses made in Cartesian and cylindrical coordinate systems are sufficient. However, in the analysis of spherical geometry structures such as spherical electric motors and brain stimulation systems, analytical calculation of self-inductance and mutual inductance coefficients in a spherical coordinate system is required. Studies conducted in spherical coordinates have become one of the important research topics in recent years. The first studies on spherical electric motors began in the early 1980s. These studies derived analytical formulas expressing the magnetic field and electric field created by a spherical coil. In a study conducted in 1975, a model with an air core and axial symmetry was addressed. In the study, they obtained a formula giving the magnetic field using the integral transformation method [9]. Another study examined the magnetic field and inductance calculations of an isolated magnetic field analytically generated by a coil having $4\pi 10^{-7}$ H/m magnetic permeability and with diameter r in the outer orbit in a linear and homogeneous isotropic environment [10].

In brain stimulation applications, an analytical method was developed in spherical coordinates to calculate the total electric field in the center caused by the coils placed in various points [11]. The analytical solution of Maxwell's equations in a spherical coordinate system is obtained based on the assumption of axial symmetry. Matute's study has revealed that the method of decomposition of variables is suitable for reaching the analytical solution for resolving some problems addressed in curvilinear coordinates that can be accepted as axially symmetrical [12]. One of the latest analytical studies conducted in the spherical coordinate system was carried out in 2019 to obtain a homogeneous magnetic field. In this study, they observed the sequential arrangement of a sizable number of coils analytically. At the end of the study, it was observed that the sequentially arranged coils with a spherical structure were more homogeneous compared to coils of equal size [13]. Most of the studies include the calculations of vector field potential **A**, magnetic field **B**, electric field **E** and self-inductance coefficient L_{ii} and mutual inductance coefficient M_{ij} based on the single-winding coil approach. Most of the analytical terms involve Legendre Polynomials or Associated Legendre Polynomials. Analytical formulas for multi-winding coil structures are not yet available in studies conducted in spherical coordinates. So, finite element analysis (FEA) is used to calculate **E**, **B,** and torque (τ) magnitudes in the applications of electrical machines with spherical geometry. Spherical induction motor studies the most important examples are [14-18]. New results for the analytical calculation of **E**, **B,** and M_{ij} sizes in spherical coordinates will contribute to shortening the calculation time and design process.

In this study; Chapter 2, describes the methods used in the general solution of maxwell's equations in spherical coordinates. Chapter 3 presents the analytical solution of field vectors originating from a single-turn magnetic-core coil (i.e. the coil is wound in a magnetic, conductive material, in this case, iron) in a spherical coordinate system. Analytical formulas were obtained for the magnitudes of the vector field potential **A**, magnetic field **B**, and electric field **E** formed by a single-turn coil with a magnetic core structure in spherical coordinates and the consistency with the analytical formulas derived for the single-turn coil with air core in the literature was examined. Chapter 4, analytical formulas were suggested expressing **A**, **E**, and **B** for a multi-winding coil structure with a magnetic core. This was the first time in the literature. By utilizing the Matlab^{TM} program, numerical values of **A** and **B** magnitudes on a plane surface were obtained under the assumption of axial symmetry in different coil structures and compared with FEA. In Chapter 5, the interaction problem of two concentric multi-winding coils is examined, and analytical formulas for self-inductance and mutual inductance coefficients are proposed. Chapter 6, an experimental setup was designed by determining the structures of two concentric coils with magnetic core and radius r_1. The mutual inductance coefficient of the coil at different angles was calculated by measuring the voltage and current values in the experimental setup. Experimental results, the FEA analysis results, and the results of analytical formulas were compared. In the study, magnetic field **B** and electric field **E** magnitudes formed by multi-winding coil structures wound on magnetic core were calculated for the first time using a spherical coordinate system. In chapter 7, the results were evaluated.it has been presented that analytical formulas can be used in the preliminary design of various spherical systems, yielding much faster results compared to Three-dimensional FEA analyses. This is important for the analysis of spherical electrical machines.

2 General Separated Solution in Spherical Coordinates

The equations determining the function of electromagnetic systems are called Maxwell's equations. The solution of Maxwell's equations under the conditions of the boundary specific to the problem makes it possible to calculate **B**, **E,** and inductance coefficients. The differential form of Maxwell's equations for electromagnetic systems which consists of linear, isotropic, permeable, and conductive materials is given below [19].

$$\nabla \bullet \mathbf{E} = \rho \qquad \text{(Gauss Law)} \qquad (1)$$

$$\nabla \bullet \mathbf{B} = 0 \qquad (2)$$

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \qquad \text{(Faraday Law)} \qquad (3)$$

$$\nabla \times \mathbf{B} = \mu \mathbf{J} + \mu \grave{o} \frac{\partial \mathbf{E}}{\partial t} \qquad \text{(Maxwell Modified Ampère Law)} \qquad (4)$$

Here, μ and $\grave{o}$, are the magnetic permeability constant and the dielectric permeability constant of the field and ρ and **J** are the free electric charge density and the current density. **E** and **B** can be expressed in terms of scalar potential (ϕ) and vector potential (**A**), as follows [20].

$$\nabla^2 \phi = 0$$

The terms of **E** and **B** are given below as the type of vector **A** [21].

$$\mathbf{E} = -i\omega \mathbf{A}$$

$$\mathbf{B} = \nabla \times \mathbf{A}$$

$$\nabla^2 \mathbf{A} - \mu \grave{o} \frac{\partial^2 \mathbf{A}}{\partial t^2} = -\mu \mathbf{J} \qquad (5)$$

Here, i and ω indicate 90° degrees phase angle and angular frequency (R/s) respectively. The solution of Eq. (5) in the spherical coordinate system provides the representation of the vector potential in spherical coordinates. Since no current flows except the r_0, θ_0 coordinates in which the coil exists, we can write $\mathbf{J} = 0$ outside the coil.

$$\nabla^2 \mathbf{A} - \mu \grave{o} \frac{\partial^2 \mathbf{A}}{\partial t^2} = 0 \qquad (6)$$

The homogeneous partial differential equation given by Eq. (6) is the electromagnetic wave equation, and **A** in linear systems with time oscillations can be written in the form below [12].

$$\mathbf{A}(r,\theta,t) = \mathbf{A}_{\varphi}(r,\theta)e^{-i\omega t} \tag{7}$$

If **A** as the term of vector field potential given by Eq. (7) is written in Eq. (6), the vector Helmholtz equation is found. The vector Helmholtz equation allows us to find the solution to the problem based on the boundary conditions regardless of time. If k^2<<1 when solving the Helmholtz equation, the term $k^2 = \omega\mu\sigma$ is ignored.

$$\nabla^2 \mathbf{A} = 0 \tag{8}$$

If $\mathbf{A_0} = A_0(r,\theta)\mathbf{e}_{\varphi}$ based on the assumption of axial symmetry, then the expanded form of Eq. (8) is found as Eq. (9) in spherical coordinates.

$$\nabla^2 \mathbf{A}\varphi = \frac{1}{r}\frac{\partial^2 (rA_{\varphi})}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2 A_{\varphi}}{\partial \theta^2} + \frac{\cot\theta}{r^2}\frac{\partial A_{\varphi}}{\partial \theta} - \frac{1}{r^2}\frac{A_{\varphi}}{\sin^2\theta} \tag{9}$$

If the partial differential equation given by Eq. (9) is solved with the method of the decomposition of the variables, the serial solution of **A** is obtained. Here, the term $P_n^1(\cos\theta)$ represents the 1st order Associated Legendre Polynomial. The equation given by Eq. (10) is expressed as the general separated solution in the spherical coordinate system [22] and so on.

$$A_{\varphi} = \sum_{1}^{\infty}(a_n r^n + b_n r^{-n-1})P_n^1(\cos\theta) \tag{10}$$

3 Calculation of Magnetic Field and Electric Field for a Single-Winding Coil with Magnetic Core

In spherical electrical machines, rotor and stator windings are wound around a magnetic core. So, analyzes must be made in the spherical coordinate system. A spherical electric machine in its simplest form consists of a single winding coil structure wound on a magnetic core. In real applications, there are multi-winding coil structures on the magnetic core. For the first problem, the definitions and geometric features required in the solution of the problem are defined in spherical coordinates and shown in Figure 1.

Figure 1a shows the magnetic sphere and the coil winding around it. The coil winding is wound outside the sphere structure and it is symmetrical concerning the z-axis. In Figure 1b, the coil structure is placed on the x-y plane. The coil is placed in the r_0 and θ_0 coordinates in the spherical coordinates. By monitoring the electric and magnetic field components of the coil on three different fields of A_0, A_1 and A_2, the representation of the vector field potential within these fields will be calculated.

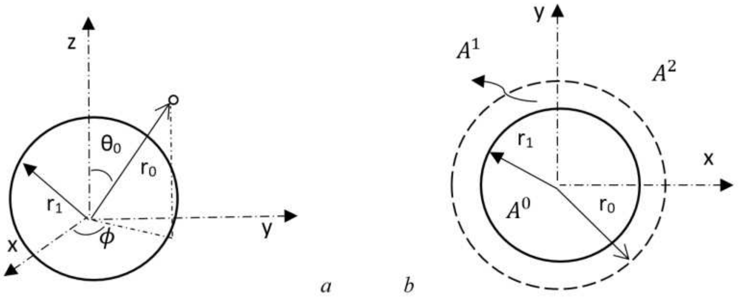


Figure 1
The cross-section view of the sphere and the winding (a)
The top view of the sphere and the winding (b)

$A^{(0)}$, $A^{(1)}$, and $A^{(2)}$ are the spherical cores, the gap regions inside and outside the radius r_0 in which the coil is wound, respectively. Where r and θ are the coordinates of the calculated point, r_1 is the radius of the spherical core geometry, r_0 and θ_0 are the coil coordinates, μ_1 and μ_0 are the magnetic permeability coefficients of the calculation points, and $\mu=\mu_1/\mu_0$ is the respective magnetic permeability coefficient. Consider region $A^{(0)}$ ($r<r_1$), the second term in Eq. (10) is neglected because it is divergent. Consider region $A^{(1)}$ ($r_1<r\leq r_0$), both two terms are used together. Consider region $A^{(2)}$ ($r\geq r_0$), the first term in Eq. (10) is neglected because it is divergent. In this case, **A** is written separately for each of the three regions as follows;

$$A^{(0)}(r,\theta)=\sum_{1}^{\infty}a_n r^n P_n^1(\cos\theta)$$

$$A^{(1)}(r,\theta)=\sum_{1}^{\infty}(b_n r^n+c_n r^{-n-1})P_n^1(\cos\theta) \tag{11}$$

$$A^{(2)}(r,\theta)=\sum_{1}^{\infty}d_n r^{-n-1}P_n^1(\cos\theta)$$

Maxwell's equations are valid for all points where physical parameters are constant. However, in some cases, physical parameters may vary such as geometry, current, μ, ò etc. In these cases, the boundary conditions given in Eq. (12) become valid [23].

$$\begin{aligned}
&\mathbf{n}\times(\mathbf{E}_2-\mathbf{E}_1)=0\\
&\mathbf{n}\times(\mathbf{H}_2-\mathbf{H}_1)=\mathbf{J}\\
&(\mathbf{B}_2-\mathbf{B}_1)\cdot\mathbf{n}=0\\
&(\mathbf{D}_2-\mathbf{D}_1)\cdot\mathbf{n}=\rho_s
\end{aligned} \tag{12}$$

Here **D** is the electric current density and ρ_s is the surface charge density, if the conductivity values of the two substances are finite, the surface current density **J**=0, and equations can be rearranged.

$$\mathbf{n}\times(\mathbf{H}_2-\mathbf{H}_1)=0 \tag{13}$$

In the coordinates of $r=r_1$ and $r=r_0$, there is no electric field change in the direction of the surface normal. Since no current occurs on the sphere surface, J=0. On the surface where the winding is ($\theta=\theta_0\ r=r_0$), **J** is non-zero. If it is solved along with the representative solutions given by Eq. (1), four equations with four unknowns are obtained (14).

$$\begin{aligned}
&(b_n-a_n)r_1^n+c_nr_1^{-n-1}=0\\
&b_nr_0^n+(c_n+d_n)r_1^n+c_nr_1^{-n-1}=0\\
&(n+1)(\frac{1}{\mu}a_n-b_n)r_1^{n-1}+nc_nr_1^{-n-2}=0\\
&(n+1)b_nr_0^n+n(d_n-c_n)r_0^{-n-1}-\frac{2n+1}{2n(n+1)}I\mu_0P_n^1(\cos\theta_0)\cos\theta_0=0
\end{aligned} \tag{14}$$

The obtained common solution to equations of Eq. (14), gives the coefficients of a_n, b_n, c_n, and d_n presented in Eq. (15). Using Eq. (11) and Eq. (15), **A** created by a single-winding coil that is wound on the magnetic core can be calculated.

$$\begin{aligned}
&a_n=\frac{(2n+1)\mu}{2n(n+1)(1+n+n\mu)}I\mu_0r_0^{-n}\sin\theta_0P_n^1(\cos\theta_0)\\
&b_n=\frac{1}{2n(n+1)}I\mu_0r_0^{-n}\sin\theta_0P_n^1(\cos\theta_0),\\
&c_n=\frac{(\mu-1)}{2n(1+n+n\mu)}I\mu_0r_0^{-n}r_1^{1+2n}\sin\theta_0P_n^1(\cos\theta_0)\\
&d_n=\frac{I\mu_0\sin\theta_0P_n^1(\cos\theta_0)r_0^{-n}}{2n(n+1)(1+n+n\mu)}((1+n+n\mu)r_0^{1+2n}+(1+n)(\mu-1)r_1^{1+2n})
\end{aligned} \tag{15}$$

E and **B** is calculated with the help of Eq. (16) depending on **A**. The expanded form of **B** is presented in Eq. (17).

$$\begin{aligned}
&\mathbf{E} = -i\omega\mathbf{A} \\
&\mathbf{B} = \nabla\times\mathbf{A} \qquad\qquad (16)\\
&\mathbf{B} = \frac{1}{r\sin\theta}\frac{\partial}{\partial\theta}(A_\varphi\sin\theta)\mathbf{e_r} - \frac{1}{r}\frac{\partial}{\partial r}(rA_\varphi)\mathbf{e_0}
\end{aligned}$$

$$\begin{aligned}
&\mathbf{B}^{(0)} = \sum_1^\infty a_n r^{n-1} n(n+1)P_n(\cos\theta)\mathbf{e}_r - \sum_1^\infty (n+1)a_n r^{n-1}P_n^1(\cos\theta)\mathbf{e}_\theta \\
&\mathbf{B}^{(1)} = \sum_1^\infty (b_n r^{n-1} + c_n r^{-n-2}) n(n+1)P_n(\cos\theta)\mathbf{e}_r - \sum_1^\infty ((n+1)b_n r^{n-1} - nc_n r^{-n-2})P_n^1(\cos\theta)\mathbf{e}_\theta \\
&\mathbf{B}^{(2)} = \sum_1^\infty n(n+1)d_n r^{-n-2}P_n(\cos\theta)\mathbf{e}_r + \sum_1^\infty nd_n r^{-n-2}P_n^1(\cos\theta)\mathbf{e}_\theta
\end{aligned}$$

(17)

Eq. (17) gives the general solution for the core structures with different magnetic permeability. In Eq. (17) the particular case of μ=1 gives the particular case of the analysis of a coil structure with an air core in spherical coordinates. When Eq. (16) and Eq. (17) are analyzed, it is seen that they are compatible with the equations given in the literature [22].

4 Calculation of the Integral Form for a Multi-Winding Coil with a Magnetic Core

By Eq. (17), the analytical form of **B** is presented for a single-winding coil wound on a sphere having a magnetic permeability of μ_1 and a radius of r_1. However, in practice, coil windings are wound in multiple windings to include a certain region (Figure 2). For the regions of A_0, A_1 and A_2, the integral form of **A** formed by a multi-winding coil is given in Eq. (18). Where r_a, r_b, θ_a, and θ_b are the starting and ending coordinates of the coil winding, N is the number of the coil winding, and S refers to the surface field of the coil winding (Figure 2).

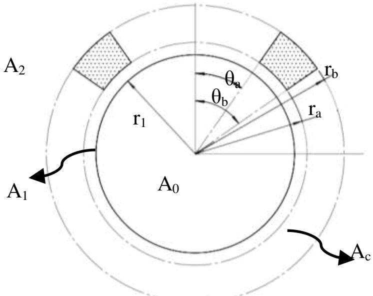


Figure 2
The structure of the multi-winding coil

$$A_i^{(0)}(r,\theta)=\int_{r_a}^{r_b}\int_{\theta_a}^{\theta_b}A^{(0)}(r,\theta)r_0dr_0d\theta_0$$

$$A_i^{(1)}(r,\theta)=\int_{r_a}^{r_b}\int_{\theta_a}^{\theta_b}A^{(1)}(r,\theta)r_0dr_0d\theta_0 \tag{18}$$

$$A_i^{(2)}(r,\theta)=\int_{r_a}^{r_b}\int_{\theta_a}^{\theta_b}A^{(2)}(r,\theta)r_0dr_0d\theta_0$$

$$S=\frac{1}{2}(r_b{}^2-r_a{}^2)(\theta_b-\theta_a)$$

$$R(u,v)=\left\{\begin{matrix} n=1, & v-u \\ n=2, & \ln(\frac{v}{u}) \\ n>2, & \frac{1}{2-n}(v^{2-n}-u^{2-n}) \end{matrix}\right\}$$

By the principle of superposition, if the integral is taken on the coefficients of a$_n$, b$_n$, c$_n$, and d$_n$ given in Eq. (15), and on the boundaries of r$_a$, r$_b$, θ$_a$, and θ$_b$ the coefficients representing the structure of a multi-winding coil is obtained.

$$a_n^i=\frac{N(2n+1)\mu I\mu_0}{S2n(n+1)(1+n+n\mu)}R(r_a,r_b)\int_{\theta_a}^{\theta_b}\sin\theta_0P_n^1(\cos\theta_0)d\theta_0$$

$$b_n^i=\frac{NI\mu_0}{S2n(n+1)}R(r_a,r_b)\int_{\theta_a}^{\theta_b}\sin\theta_0P_n^1(\cos\theta_0)d\theta_0$$

$$c_n^i=\frac{N(\mu-1)I\mu_0}{S2n(1+n+n\mu)}r_1^{1+2n}R(r_a,r_b)\int_{\theta_a}^{\theta_b}\sin\theta_0P_n^1(\cos\theta_0)d\theta_0$$

$$d_n^i=\frac{N(\mu-1)I\mu_0}{S2n(n+1)(1+n+n\mu)}\left(\begin{matrix}\left(\frac{(1+n+n\mu)}{n+3}\left(r_b^{3+n}-r_a^{3+n}\right)\right)\\ +(1+n)(\mu-1)r_1^{1+2n}R(r_a,r_b)\end{matrix}\right)\int_{\theta_a}^{\theta_b}\sin\theta_0P_n^1(\cos\theta_0)d\theta_0 \tag{19}$$

Assume that the region with the coil winding is divided into two parts from the radius of r$_0$ to calculate the effect of multi-winding in the region of A$_c$ ($r_a \leq$ r $\leq r_b$). The coils that need to be calculated in the range of r_a < r and in the A$_c$(r,θ) region, perform as in the A$_2$ region, and the coils in the range of r≤ r_b perform as in the A$_1$ region. In this case, the total vector field potential formed by two regions helps to calculate the magnitudes of **A**, **E**, and **B** approximately for this region. This assumption gives better results in case the value of □$r = r_b - r_a$ is smaller.

$$A_i^{(c)}(r,\theta) = \int_{r_a}^{r}\int_{\theta_a}^{\theta_b} A^{(2)}(r,\theta) r_0 dr_0 d\theta_0 + \int_{r}^{r_b}\int_{\theta_a}^{\theta_b} A^{(1)}(r,\theta) r_0 dr_0 d\theta_0$$

$$b_n^c = \frac{NI\mu_0}{S2n(n+1)} R(r,r_a) \int_{\theta_a}^{\theta_b} \sin\theta_0 P_n^1(\cos\theta_0) d\theta_0$$

$$c_n^c = \frac{N(\mu-1)I\mu_0}{S2n(1+n+n\mu)} r_1^{1+2n} R(r,r_a) \int_{\theta_a}^{\theta_b} \sin\theta_0 P_n^1(\cos\theta_0) d\theta_0$$

$$d_n^c = \frac{N(\mu-1)I\mu_0}{S2n(n+1)(1+n+n\mu)} \left(\begin{array}{l} \left(\dfrac{(1+n+n\mu)}{n+3}\left(r_b^{3+n} - r^{3+n}\right)\right) \\ +(1+n)(\mu-1) r_1^{1+2n} R(r,r_a) \end{array} \right) \int_{\theta_a}^{\theta_b} \sin\theta_0 P_n^1(\cos\theta_0) d\theta_0 \tag{20}$$

$$\begin{aligned}
A_i^{(0)}(r,\theta) &= \sum_1^\infty a_n^i r^n P_n^1(\cos\theta) \\
A_i^{(1)}(r,\theta) &= \sum_1^\infty (b_n^i r^n + c_n^i r^{-n-1}) P_n^1(\cos\theta) \\
A_i^{(2)}(r,\theta) &= \sum_1^\infty d_n^i r^{-n-1} P_n^1(\cos\theta) \\
A_i^{(c)}(r,\theta) &= \sum_1^\infty (b_n^c r^n + (c_n^c + d_n^c) r^{-n-1}) P_n^1(\cos\theta)
\end{aligned} \tag{21}$$

$$\begin{aligned}
\mathbf{B}_i^{(0)} &= \sum_1^\infty a_n^i r^{n-1} n(n+1) P_n(\cos\theta)\mathbf{e}_r - \sum_1^\infty (n+1) a_n^i r^{n-1} P_n^1(\cos\theta)\mathbf{e}_\theta \\
\mathbf{B}_i^{(1)} &= \sum_1^\infty (b_n^i r^{n-1} + c_n^i r^{-n-2}) n(n+1) P_n(\cos\theta)\mathbf{e}_r \\
&\quad - \sum_1^\infty ((n+1) b_n^i r^{n-1} - n c_n^i r^{-n-2}) P_n^1(\cos\theta)\mathbf{e}_\theta \\
\mathbf{B}_i^{(2)} &= \sum_1^\infty n(n+1) d_n^i r^{-n-2} P_n(\cos\theta)\mathbf{e}_r + \sum_1^\infty n d_n^i r^{-n-2} P_n^1(\cos\theta)\mathbf{e}_\theta \\
\mathbf{B}_i^{(c)} &= \sum_1^\infty (b_n^c r^{n-1} + (c_n^c + d_n^c) r^{-n-2}) n(n+1) P_n(\cos\theta)\mathbf{e}_r \\
& - \sum_1^\infty ((n+1) b_n^c r^{n-1} - n(c_n^c - d_n^c) r^{-n-2}) P_n^1(\cos\theta)\mathbf{e}_\theta
\end{aligned} \tag{22}$$

The coefficients presented in Eq. (19) and Eq. (20) are used to calculate the magnitudes of **A** and **B** given in Eq. (21) and Eq. (22). Figure 3 shows how to calculate A and B. With the help of the Matlab program, the analysis parameters given in Table 1 are calculated using the analytical formulas given in Eq. (21) and Eq. (22). The obtained numerical results are given in Figure 4 and Figure 5.

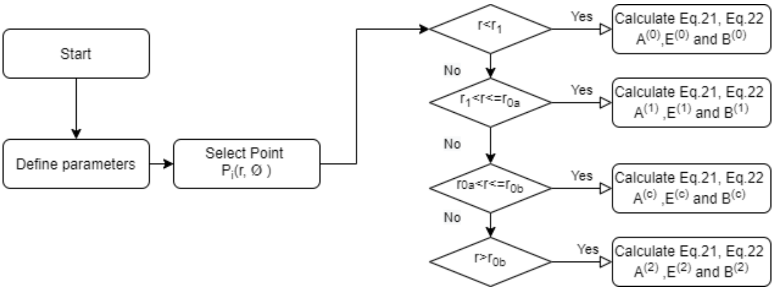


Figure 3

A, E and **B** calculation flowchart

Table 1

The analysis parameters

Parameter	Value	Parameter	Value
I	1 A	θ_0	45°
$\mu_0(H/m)$	$4\pi 10^{-7}$	$\Delta\theta$	10°
$\mu(H/m)$	1, 10, 100	θ	22.5°,30°,45°
$r_0(mm)$	34	N	100
$r_1(mm)$	30		
$r(mm)$	$0<r<50$		

Figure 4 shows the variation with μ of **B** and **A** on a sphere of radius r_1 of a coil placed in r_a, r_b, θ_a, and θ_b coordinates. The values of **A** in regions far from the coil winding have small and smooth transitions. Where θ=45° is the direction in which the magnetic effects are the highest. Thus, the graphic in Figure 4f is the one that the magnetic effect can be observed most clearly. It is obvious that as μ gets larger, magnetic effects accumulate on the surface of the sphere.

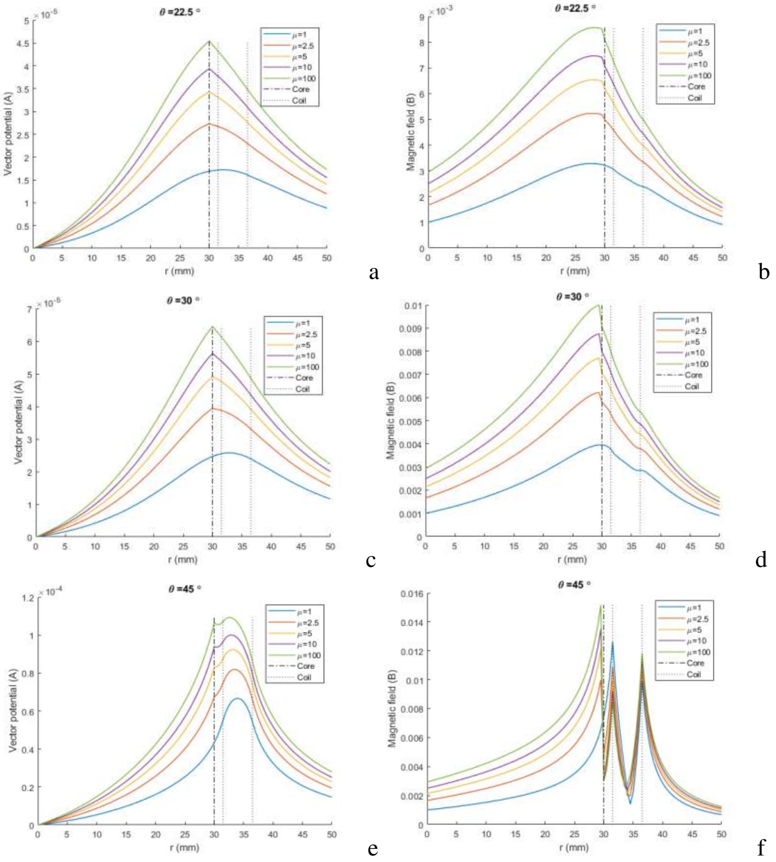


Figure 4

For θ=22.5° the graphic view of the variation of **A** for different μ values (a). For θ=22.5° the graphic view of the variation of **B** for different μ values (b). For θ=30° the graphic view of the variation of **A** for different μ values (c). For θ=30° the graphic view of the variation of **B** for different μ values (d). For θ=45° the graphic view of the variation of **A** for different μ values (e). For θ=45° the graphic view of the variation of **B** for different μ values (f).

Figure 5, shows the variations of **B** and **A** on the x-z plane with the different μ values. When Figure 4 and Figure 5 are examined, it is witnessed that as μ increases, the magnitudes of the magnetic field and the vector potential come closer to the sphere surface.

The coil and sphere geometries, the features of which given in Table 1 with the ANSYS Maxwell program were formed under the assumption of 2D axial symmetry. As the magnetostatic analysis, three different analyses were performed as μ=1, μ=10, and μ=100.

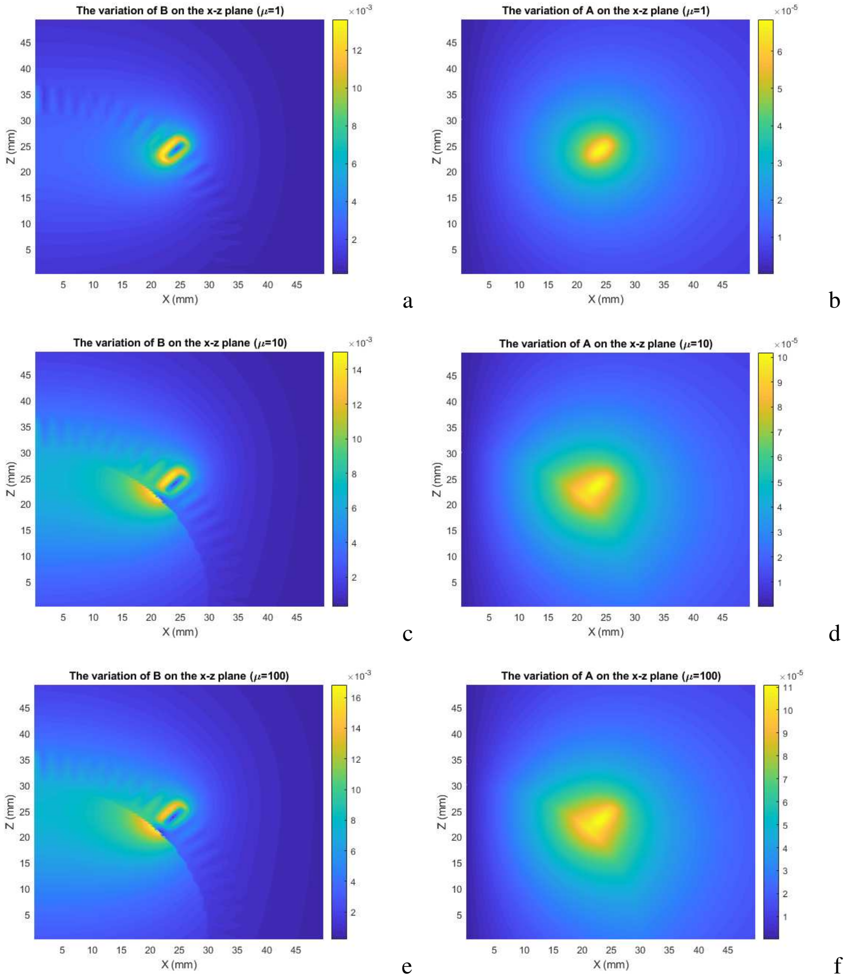


Figure 5

For μ=1, the variation of **B** on the x-z plane (a). For μ=1, the variation of **A** on the x-z plane (b). For μ=10, the variation of **B** on the x-z plane (c). For μ=10, the variation of **A** on the x-z plane (d). For μ=100, the variation of **B** on the x-z plane (e). For μ=100, the variation of **A** on the x-z plane (f).

The distribution of the magnitude of the magnetic field **B** on the x-z plane which was obtained with the ANSYS Maxwell program is presented in Figure 6. When Figure 6 and Figure 5 are observed, it is seen that the FEA results and the distribution of the magnetic field obtained from analytical calculations give similar results.

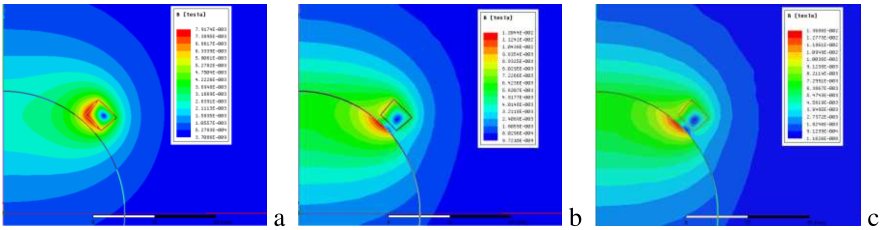
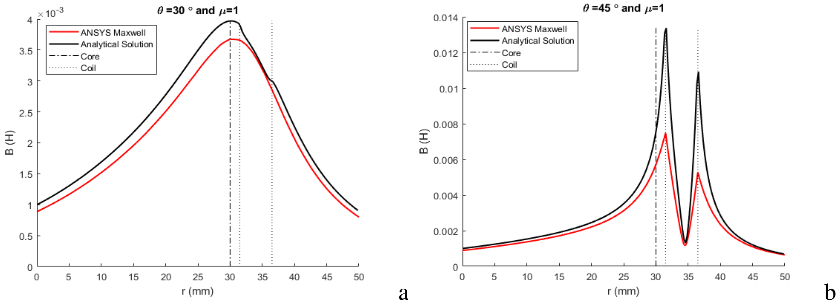


Figure 6
For ANSYS Maxwell μ=1, the distribution of the magnetic field on the x-z plane (a). For ANSYS Maxwell μ=10, the distribution of the magnetic field on the x-z plane (b). For ANSYS Maxwell μ=100, the distribution of the magnetic field on the x-z plane (c).

In Figure 7 for θ=30° and θ=45° the variation of **B** on r is compared with the FEA results. When the results are assessed, it is seen that the magnitudes of the magnetic field obtained with analytical formulas are consistent with the ones obtained from the model of the FEA. The differences in the results become apparent around the coil geometry. In this case, in the calculation of multi-winding coil results, the formula used for the region A_c among the coil windings is valid for small values of r_b-r_a. When the distance between coil geometries expands, the differences increase due to the single-coil approach.



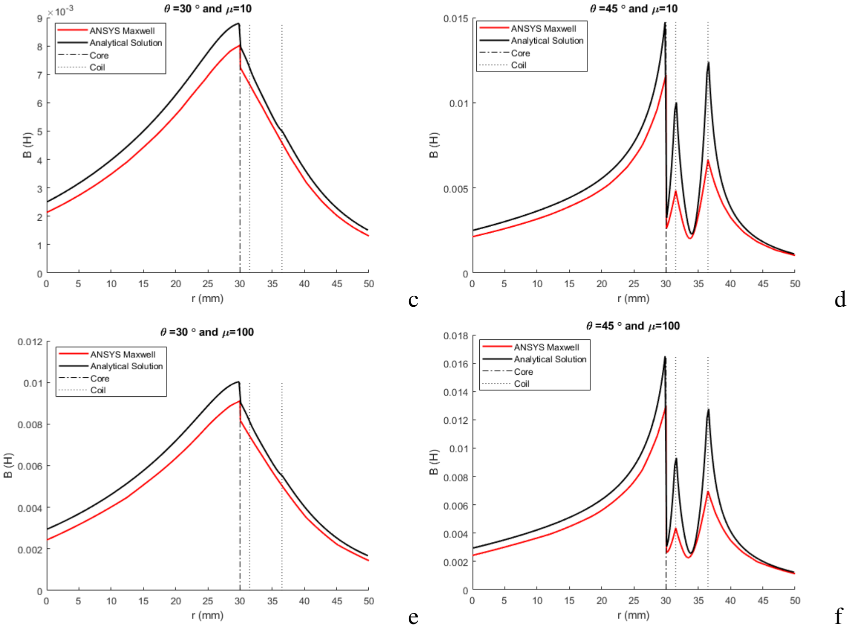


Figure 7

in the coordinate of θ=30° for μ=1, μ=10, μ=100, the comparison of the magnitudes of the magnetic field with the FEA results (a)(c)(e). In the coordinate of θ=45° for μ=1, μ=10, μ=100, the comparison of the magnitudes of the magnetic field with the FEA results (b)(d)(f).

5 Calculation of Inductance Coefficients

5.1 The Self-Inductance Coefficient

In electromagnetic systems, the variation of the magnetic field causes an electric field. The voltage is induced on a ring-shaped copper wire that is exposed to a variable magnetic field. The Faraday Law gives the relationship between the current generated in the wire exposed to a variable magnetic field and the variation of the magnetic field [19]. If the connection between **E** and **A** is used, the voltage V in the wire can be calculated applying the magnitude of **A**;

The voltage generated in the N coil winding;

$$V=-\frac{N}{S}\iint_{dS'}\left\{\int_{dr}\mathbf{E}\Box d\mathbf{r}\right\}dS'$$
$$V=i\omega\frac{N}{S}\iint_{dS'}\left\{\int_{dr}\mathbf{A}\Box d\mathbf{r}\right\}dS' \tag{23}$$

The innermost integral is the voltage formed by a single-winding coil (r_0,θ_0) on itself; Here, **E** is taken as $E(r_0,\theta_0)$. This voltage is integrated on the cross-section and multiplied by the number of windings per unit area.

$$V=i\omega\frac{N}{S}\iint_{\substack{cross\\ \sec tion}}\left\{A(r_0,\theta_0)r_0\sin\theta_0\int_0^{2\pi}d\varphi\right\}dS$$
$$V=i\omega\frac{N}{S}\iint_{\substack{cross\\ \sec tion}}\left\{A(r_0,\theta_0)r_0\sin\theta_0\int_0^{2\pi}d\varphi\right\}r_0dr_0d\theta_0 \tag{24}$$

Where **A**, $A^{(c)}$, $dS=r_0dr_0d\theta_0$ and $dr=r_0\sin\theta_0d\varphi$ on the cross-section, if the equation is arranged by substituting Eq. (24) with Eq. (21); then

$$V=\frac{i\omega I\mu_0N^2}{S^2}\iint_{dS'}\left\{\int_{d\varphi=0}^{2\pi}\sum_{n=1}^{\infty}\frac{1}{2n(n+1)}(R(r_0,r_a)r_0^{\,n+1}+(\frac{(\mu-1)(n+1)}{(1+n+n\mu)}R(r_0,r_a)r_1^{\,1+2n}r_0^{\,-n}+\frac{(\mu-1)}{(1+n+n\mu)}\left(\begin{array}{l}\left(\frac{(1+n+n\mu)}{n+3}\left(r_0^{\,-n}r_b^{\,3+n}-r_0^{\,2}\right)\right)\\ +(1+n)(\mu-1)r_0^{\,-n}r_1^{\,1+2n}R(r_0,r_a)\end{array}\right)))Q_nP_n^1(\cos\theta_0)d\varphi\right\}dS'$$

$$V=\frac{\pi i\omega I\mu_0N^2}{S^2}\int_{dr_0=r_a}^{r_b}\sum_{n=1}^{\infty}\frac{1}{n(n+1)}Q_n^{\,2}(R(r_0,r_a)r_0^{\,n+2}+\frac{(\mu-1)(n+1)}{(1+n+n\mu)}R(r_0,r_a)r_1^{\,1+2n}r_0^{\,-n+1}+\left(\begin{array}{l}\frac{(\mu-1)}{(n+3)}\left(\left(r_0^{\,-n+1}r_b^{\,3+n}-r_0^{\,4}\right)\right)\\ +\frac{1}{(1+n+n\mu)}(1+n)r_0^{\,-n+1}r_1^{\,1+2n}R(r_0,r_a)\end{array}\right))dr_0 \tag{25}$$

Let $Q_n\left(\theta_a,\theta_b\right)=\int_{\theta_a}^{\theta_b}\sin\theta_0P_n^1(\cos\theta_0)d\theta_0$ and the Z is the impedance of the coil;

$$\mathrm{Z}=\frac{\mathrm{V}}{\mathrm{I}}=\mathrm{i}\omega\mathrm{L} \tag{26}$$

$$L=\frac{\pi\mu_0N^2}{S^2}\int_{dr_0=r_a}^{r_b}\sum_{n=1}^{\infty}\frac{Q_n^{\,2}}{n(n+1)}(R(r_0,r_a)r_0^{\,n+2}+\frac{(\mu-1)(n+1)}{(1+n+n\mu)}R(r_0,r_a)r_1^{\,1+2n}r_0^{\,-n+1}+\left(\begin{array}{l}\frac{(\mu-1)}{(n+3)}\left(\left(r_0^{\,-n+1}r_b^{\,3+n}-r_0^{\,4}\right)\right)\\ +\frac{1}{(1+n+n\mu)}(1+n)r_0^{\,-n+1}r_1^{\,1+2n}R(r_0,r_a)\end{array}\right))dr_0 \tag{27}$$

Eq. (27) gives the self-inductance coefficient of a multi-winding coil. Since the integral process cannot be calculated analytically, they are calculated numerically.

5.2 The Mutual Inductance Coefficients

To calculate the mutual inductance coefficient, let's consider two concentric single-winding coils in spherical coordinates as seen in Figure 8a. When a variable current flows on coil 1, a variable magnetic field is generated in the air and magnetic core. An electromotive force (emf-voltage) is induced on Coil 2 due to the changing magnetic field. M_{ij} is used to calculate the interaction between the two coils. Here, γ is the angle between the axes of the coils, α and β is the angular coordinate of the coils concerning their set of axes, a and b are the r coordinates where the coils are located. φ' and θ' are the angular coordinates concerning the fixed set of axes of the second coil. The representation of the coils as multi-windings is given in Figure 8b and Figure 8c.

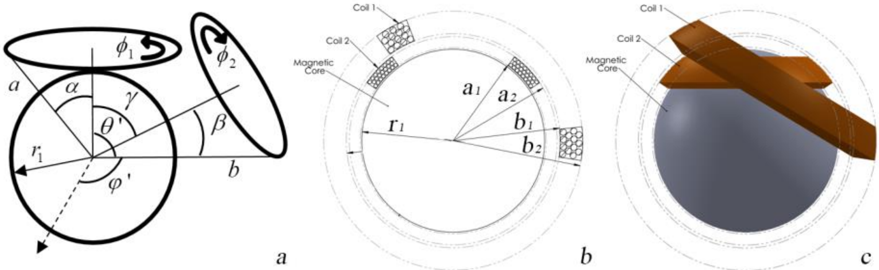


Figure 8
Two concentric single-winding coils (a), two concentric multi-winding coils (b), three-dimensional view (c)

If a<b in the single-winding coil method, the mathematical form of the voltage that the first coil generates on the second coil is given in Eq. (23). The coil with the N_1 number of windings and radius a is located in the ∝ position in spherical coordinates. The vector potential of $A_i^{(2)}(b,\theta')$ which the coil creates is given by Eq. (21).

$$A_i^{(2)}(b,\theta')=\sum_1^{\infty} d_n^i b^{-n-1} P_n^1(\cos\theta')$$

$$d_n^i=\frac{N_1(\mu-1)I\mu_0}{S_1 2n(n+1)(1+n+n\mu)}\left(\begin{matrix}\left(\dfrac{(1+n+n\mu)}{n+3}\left(a_2^{3+n}-a_1^{3+n}\right)\right)\\ +(1+n)(\mu-1)r_1^{1+2n}R(a_1,a_2)\end{matrix}\right)\int_{\alpha_1}^{\alpha_2}\sin\alpha_0 P_n^1(\cos\alpha_0)d\alpha_0 \tag{28}$$

Using Eq. (21) and Eq. (23), the mathematical form of the voltage expression generated in the second coil with the N_2 number of windings and located in b and β coordinates is given below;

$$V = \frac{N_2}{S_2}\iint_{dS'}\left\{\int_{dr} A\Box d\mathbf{r}\right\}dS'$$

$$V = i\omega\frac{N_2}{S_2}\iint_{dS'}\left\{\int_{d\varphi}^{2\pi}\sum_{n=1}^{\infty}\left(\begin{array}{l}\left(\frac{(1+n+n\mu)}{n+3}\left(a_2^{3+n}-a_1^{3+n}\right)\right)\\ +(1+n)(\mu-1)r_1^{1+2n}R(a_1,a_2)\end{array}\right)b^{-n}P_n^1(\cos\theta')\sin\beta d\varphi\right\}dS'$$

(29)

Here, if r=b and $dS' = rdrd\beta$; then

$$V = i\omega\sum_{n=1}^{\infty}\frac{\pi N_2 N_1(\mu-1)I\mu_0}{S_2 S_1 n(n+1)(1+n+n\mu)}e_n\left\{\begin{array}{ll} n=1, & b_2-b_1 \\ n=2, & \ln(\frac{b_2}{b_1}) \\ n>2, & \frac{1}{2-n}(b_2^{2-n}-b_1^{2-n}) \end{array}\right\}\beta\alpha P_n(\cos\gamma)$$

$$e_n = \left(\begin{array}{l}\left(\frac{(1+n+n\mu)}{n+3}\left(a_2^{3+n}-a_1^{3+n}\right)\right)\\ +(1+n)(\mu-1)r_1^{1+2n}R(a_1,a_2)\end{array}\right)$$

(30)

$$Z = \frac{V}{I} = i\omega M_{12}$$

$$M_{12} = \sum_{n=1}^{\infty}\frac{\pi N_2 N_1(\mu-1)\mu_0}{S_2 S_1 n(n+1)(1+n+n\mu)}e_n\left\{\begin{array}{ll} n=1, & b_2-b_1 \\ n=2, & \ln(\frac{b_2}{b_1}) \\ n>2, & \frac{1}{2-n}(b_2^{2-n}-b_1^{2-n}) \end{array}\right\}\beta\alpha P_n(\cos\gamma)$$

(31)

For a multi-winding coil, where a_1, a_2, b_1, and b_2 are the coil radius, $\alpha_1, \alpha_2, \beta_1$ and β_2 are the angular coordinates concerning their set of axes, γ is the angle between the coil axes, the mutual inductance coefficients (M_{12}) of two coils is computed with Eq. (31) (Figure 8). Figure 9 shows how to program calculate M_{ij} for each $P_i(r,\theta)$,

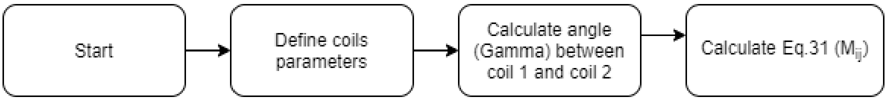

Start
Define coils parameters
Calculate angle (Gamma) between coil 1 and coil 2
Calculate Eq.31 (M$_{ij}$)

Figure 9
Mutual Inductance Coefficients

The mutual inductance coefficient is of great importance in the time-dependent analysis of wireless energy transfer and mobile electrodynamic systems. Magnetic field magnitudes in the spherical coordinate system under the assumption of 2-dimensional axial symmetry can be calculated rapidly with the programs of ANSYS Maxwell, FEA, etc. However, if there is a second coil in the geometry that breaks the axial symmetry, it becomes necessary to design and calculate the models in 3 dimensions. In this case, the FEA program requires a long, time-consuming analysis by using a large number of elements. The fact that the analytical equations are given in Eq. (31) can be calculated easily and quickly will provide the result in a short time at the preliminary design stage.

Table 2
The mutual inductance analysis parameters

Parameter	Coil 1	Coil 2
r_1	30 mm	
a_1, b_1	36.5 mm	44 mm
a_2, b_2	41.5 mm	49 mm
α_1, β_1	75 °	75 °
α_2, β_2	105 °	105 °
γ	0°	0°≤ γ ≤360°
μ	1, 10, 100, 1000	
N_1, N_2	180	150
D_1, D_2	0.75 mm	0.5 mm
R1, R2	3.35 Ω,	2.6 Ω
R3	10.4 Ω	

The results of the FEA program and analytical calculations by using the geometric properties given in Table 2 are presented in Figure 10. The problem parameters discussed in the experimental studies reviewed in Chapter 6 were also chosen the same values. It is seen that the mutual inductance results obtained from analytical calculations and the FEA model are compatible. While calculations take 60 minutes in the FEA model, calculations are completed in as little as 1 minute with analytical formulas using the suggested approach on the same computer. Therefore, the suggested approach can be prefered for the necessary calculations at the preliminary design stage rather than the FEA application.

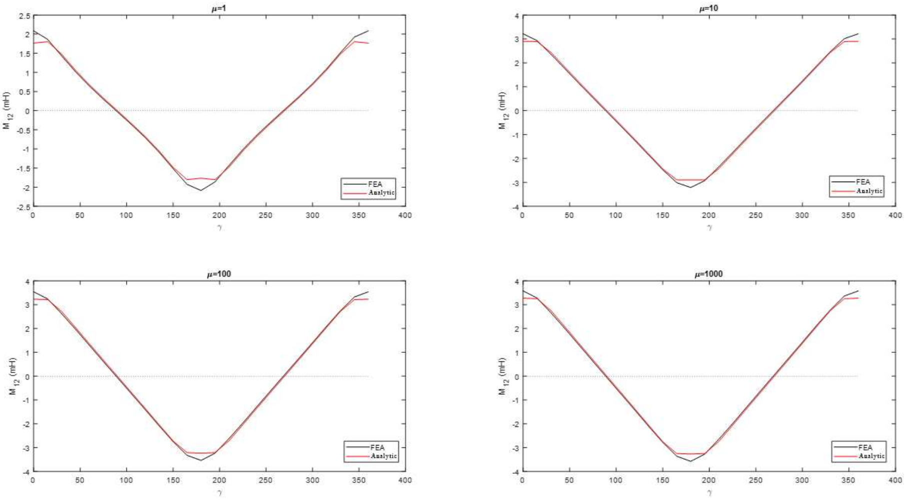


Figure 10

The comparison of the FEA and analytical calculations at different values of μ

6 Experimental Results

In order to compare the analytical formulas found for mutual inductance and self-inductance coefficients with the experimental results, an experimental setup that can be produced with a 3D printer was designed (Figure 11). The experiment setup consists of a fixed coil (Coil 1), a moving coil (Coil 2), a magnetic core and carrier legs. The descriptions of the items of the experimental setup are given in Table 3. The design parameters are given in Table 2.

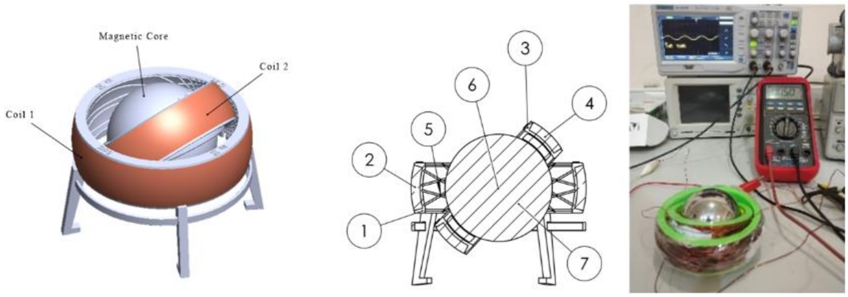


Figure 11

the Experiment Setup

Table 3

The descriptions of the experimental items

Number	Description
1	The body part on which the fixed coil is wound
2	The windings of the fixed coil
3	The body part on which the moving coil is wound
4	The windings of the moving coil
5	Angular measuring lines marked at an angle of 15° on the fixed body to measure at different angles.
6	The central rotation point
7	The magnetic Sphere

After the body parts are produced with the 3D printer, 180 turns of Ø0.75 mm diameter of copper wire are wound on the body (Coil 1), and 150 turns of Ø0.5 mm diameter of copper wire are wound on the second body (Coil 2). By modeling the coil structures in the ANSYS Maxwell program, the self-inductance coefficients are compared with the experimental results in Table 4. It was observed that the values of the inductance coefficients in all three studies were quite compatible with each other. The devices and the models of which used in the experiment are given in Table 5.

Table 4

The comparison of self-inductance coefficients

Self-inductance	Coil 2	Coil 1
Measurement	4,26 mH	2,86 mH
The Maxwell program	5,196 mH	3,729 mH
Analytical results	4.662 mH	3.334 mH

Table 5

The devices used in measurements and their features

Device Type	Brand/Model
Oscilloscope	AA TECH/ADS-3072B Digital Storage Oscilloscope
	Gw INSTEK/GDS-1022 DSO
Power Unit	Gw INSTEK/SFG-2107 Function Generator
Inductance Meter	UNI-T/UT600
Multi-meters	BRYMEN BM510, FLUKE 106

In the experiment, L_{ii} was measured with the UNI-T measuring device. Since it was not possible to measure M_{ij} directly, the current and voltage values in Coil 1 and Coil 2 were measured and calculated with Eq. (32) and Eq. (33).

$$\mathrm{I} = \mathrm{I}_{\max} \cos(\mathrm{wt}) \tag{32}$$

$$e_2 = M_{12} \frac{dI_1}{dt} \tag{33}$$

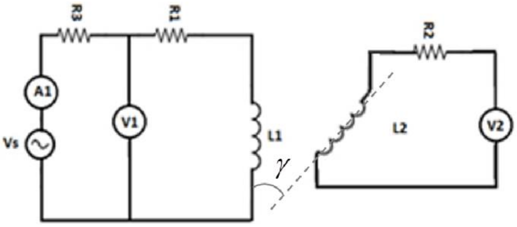


Figure 12
Electrical circuit diagram

Figure 12 shows the experimental setup circuit. The fixed coil is connected in series with a 50 Hz, 12 V of power supply with a 10.4 Ω of resistance, and the current and voltage on the fixed coil were recorded with the voltmeter (V1) and ammeter (A1). R1 and R2 are due to coil wires. R3 is used for limiting current. The voltage value formed by connecting the moving coil leads to the voltmeter (V2) was recorded in the range of 0°-360° by changing the γ angle at 15° intervals.

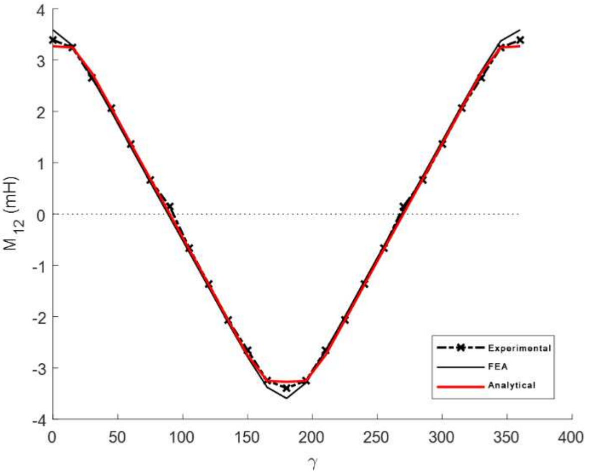


Figure 13
M_{12} at different γ values, the comparison of experimental results, analytical results and FEA results

Based on the equivalent circuit approach from the gathered data, M_{12} was calculated using Eq. (31). The comparison of the data obtained from the experimental results with the analytical formulas and the FEA results is presented in Figure 13. It is evident that the analytical calculations made for M_{12} are consistent with the experimental results and the FEA results.

$$RMSE = \sqrt{\frac{1}{D}\sum_{n=1}^{D}(y_n - \overline{y}_n)^2} \tag{34}$$

Mean square error (MSE) [24] and root mean square error (RMSE) [25-27] are common methods use to compare the results of actual measurement values and calculation values. D is the number of elements, y_n is the measured values, and $\overline{y}_n$ is the located values. So, the analytical and FEA RMSE values are calculated by Eq. (34) of 0.0714 and 0.1073, respectively.

Conclusions

In this study, the terms of **B** and **E** were calculated analytically by forming the geometry of a multi-winding coil with a magnetic core in spherical coordinates. Analytical formulas were derived for the multi-winding coil approach, given in Eq. (22). Since there are no existing analytical terms referring to the coils with the magnetic core in spherical coordinates, the obtained results were compared with the results of an air-core coil in spherical coordinates in the literature. The results were found to be consistent with the literature [22]. In order to view the effect of the magnetic sphere, different μ values and magnetic field distribution in different directions were examined. As a result of the analyses, it was observed that as the μ value increases, **B** accumulates towards the surface of the sphere (Figure 4). The coil structure with geometric properties is given in Table 1, the magnitudes of which **B** and **A** were obtained by using the formulas given in Eq. (22), and the FEA model was designed under the assumption of axial symmetry. When the results were examined, it was seen that the results of the analytical model and the FEA model were consistent (Figure 5 and Figure 6).

The experimental setup given in Figure 11 was created as the analytical and the FEA model. L_{ii} for both coils in the experiment (Table 5) and M_{ij} at different γ angles were computed by analytical and FEA models. It was observed that the numerical results were consistent with the experimental results (Figure 13). When coils are intertwined, a small collapse is detected in the analytical solution. This collapse is caused by some of the windings that are in the reverse direction during the intertwining of the coil windings. However, the results gathered by analytical calculations are quite compatible with the experimental results. It takes 15 minutes on an average computer to create a three-dimensional FEA model and calculate the mutual inductance coefficient at an angle of 15°. The same results are measured in only 15 seconds using the developed analytical method. The obtained analytical results and their computation times are considered, it is seen that analytical results are useful. The FEA results are considered, despite the small deviations between the analytical solution and the experimental results is seen that the results obtained by the analytical method are acceptable. Considering the high cost of the FEA programs and long computation times in three-dimensional models, the importance of using the suggested analytical solution is evident as a fast and free design tool in scientific studies.

References

[1] Conway, J. T. (2011) Mutual inductance for an explicitly finite number of turns. Progress In Electromagnetics Research B, 28, 273-287

[2] Ravaud, R., Lemarquand, G., Lemarquand, V., Babic, S., & Akyel, C. (2010) Mutual inductance and force exerted between thick coils. Progress In Electromagnetics Research, 102, 367-380

[3] Babic, S. I., & Akyel, C. (2006) New analytic-numerical solutions for the mutual inductance of two coaxial circular coils with rectangular cross section in air. IEEE Transactions on Magnetics, 42(6), 1661-1669

[4] Conway, J. T. (2008) Noncoaxial inductance calculations without the vector potential for axisymmetric coils and planar coils. IEEE Transactions on Magnetics, 44(4), 453-462

[5] Conway, J. T. (2013) Analytical solutions for the self-and mutual inductances of concentric coplanar disk coils. IEEE transactions on magnetics, 49(3), 1135-1142

[6] Yang, T. T., & Yang, J. J. (1975) The Effect of Cylindrical Ferromagnetic Shells on the Self and Mutual Inductance of Parallel Wires. IEEE Transactions on Electromagnetic Compatibility, (4), 234-237

[7] Babic, S. I., & Akyel, C. (2008) Calculating mutual inductance between circular coils with inclined axes in air. IEEE Transactions on Magnetics, 44(7), 1743-1750

[8] Conway, J. T. (2012) Exact solutions for the mutual inductance of circular coils and elliptic coils. IEEE transactions on magnetics, 48(1), 81-94

[9] Lipinski, W., Rolicz, P., & Sikora, R. (1975) Application of integral transforms to the analysis of the magnetic field of a spherical coil. IEEE Transactions on Magnetics, 11(5), 1552-1554

[10] Semenov, V. G. (1990) Synthesis of spherical methods of determining magnetic field source parameters of internal and external origin. Measurement Techniques, 33(12), 1236-1240

[11] Eaton, H. (1992) Electric field induced in a spherical volume conductor from arbitrary coils: application to magnetic stimulation and MEG. Medical and Biological Engineering and Computing, 30(4), 433-440

[12] Matute, E. A. (2005) On The Vector Solutions Of Maxwell Equations In Spherical Coordinate Systems, Rev. Mex. Fis. , Vol. E 51, 31-36

[13] Liu, C. Y., Andalib, T., Ostapchuk, D. C. M., & Bidinosti, C. P. (2020) Analytic models of magnetically enclosed spherical and solenoidal coils. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 949, 162837

[14] Lee, K. M., Son, H., & Joni, J. (2005, April) Concept development and design of a spherical wheel motor (SWM). In Proceedings of the 2005

IEEE International Conference on Robotics and Automation (pp. 3652-3657) IEEE

[15] Dehez, B., Galary, G., Grenier, D., & Raucent, B. (2006) Development of a spherical induction motor with two degrees of freedom. IEEE Transactions on Magnetics, 42(8), 2077-2089

[16] Fernandes, J. F., & Branco, P. C. (2016) The shell-like spherical induction motor for low-speed traction: electromagnetic design, analysis, and experimental tests. IEEE transactions on industrial electronics, 63(7), 4325-4335

[17] Zhang, C., Yuan, L., Zhang, J., Chen, J., Chen, C. Y., Chen, S., & Yang, G. (2018, April) Analytical models of electromagnetic field and torques in a novel reaction sphere actuator. In 2018 IEEE International Conference on Applied System Invention (ICASI) (pp. 271-274) IEEE

[18] Zhang, J., Yuan, L., Liao, Y., Zhang, C., Chen, C. Y., Chen, S., & Yang, G. (2018, April) Torque optimization of a novel reaction sphere actuator based on support vector machines. In 2018 IEEE International Conference on Applied System Invention (ICASI) (pp. 263-266) IEEE

[19] Griffiths, D. J. (1998) Introduction to electrodynamics. New Jersey: Prentice Hall., 3th ed., ISBN 0-13-805326-X

[20] Jackson, J. D. (1962) Classical electrodynamics john wiley & sons. Inc., New York, 13, Chapter 5, LCCCN:62-8774

[21] Paul, C. R. (2010) Inductance: loop and partial. John Wiley & Sons. New Jersey A.B.D. Chapter 3, ISBN 978-0-470-46188-4

[22] Smythe, W. B. (1989) Static and dynamic electricity. Taylor & Francis Publisher New York A.B.D., ISBN 0-89116-916-4

[23] Theodoulidis, T., & Kriezis, E. E. (2006) Eddy Current Canonical Problems (with Applications to Nondestructive Evaluation) Tech Science Press. Forsyth, GA, USA., 978-0971788015

[24] Hedrea, E. L., Precup, R. E., Roman, R. C., & Petriu, E. M. (2021) Tensor product-based model transformation approach to tower crane systems modeling. Asian Journal of Control., 23 (3), 1313-1323

[25] Precup, R. E., Teban, T. A., de Oliveira, T. E. A., & Petriu, E. M. (2016, July) Evolving fuzzy models for myoelectric-based control of a prosthetic hand. In 2016 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (pp. 72-77) IEEE

[26] Juang, C. F., Lin, Y. Y., & Huang, R. B. (2011) Dynamic system modeling using a recurrent interval-valued fuzzy neural network and its hardware implementation. Fuzzy sets and systems, 179(1) 83-99

[27] Precup, R. E., Teban, T. A., Albu, A., Borlea, A. B., Zamfirache, I. A., & Petriu, E. M. (2019, June) Evolving fuzzy models for prosthetic hand myoelectric-based control using weighted recursive least squares algorithm for identification. In 2019 IEEE International Symposium on Robotic and Sensors Environments (ROSE) (pp. 164-169) IEEE

Multi-Component Statistical Research in the area of Half-Hard Cast Iron Roll Manufacturing

Imre Kiss

University Politehnica Timișoara, Faculty of Engineering Hunedoara
Department of Engineering & Management
5, Revolutiei, Hunedoara, Romania
e-mail: imre.kiss@fih.upt.ro

Abstract: The current analysis is based on the concept that the proper quality of a particular type of alloy, such as half-hard cast irons and their properties, are determined by chemical composition and a proper melting and alloying processing, as well as, a special nodulizing treatment, assuring the graphite's nodular form. This analysis follows several key aspects of the manufacturing of half-hard cast iron rolls (also called "ductile iron rolls"), using the multivariate statistical research used as modelling approach upon the industrial data. In this sense, several results of a complex study on the half-hard cast iron rolls are presented, regarding the cumulative influences of several chemical components of the half-hard cast iron (Phosphorus, Sulphur and Magnesium), upon the Hardness, which is the common method of testing rolls, for the quality and predicted wear properties. The performed research herein has generated a number of multi-component regression equations and correlation coefficients, determined to the 3rd and 4th dimension spaces. Also generated are several regression surfaces and correlative level curves, which define proper technological areas. For the multiple regression equations and for the graphical addenda the Matlab software was used.

Keywords: cast iron rolls manufacturing; hardness; Phosphorus; Sulphur; Magnesium; multivariate regression analysis; regression equation; correlation charts

1 Introduction

In grey cast iron, graphite is present in the form of flakes [1-8]. Each of these graphite flakes under the concentrated action of an important effort, can cause the formation of cracks. In cast iron with spheroidal (nodular) graphite, known as ductile cast iron, graphite is no longer arranged in these flakes, but crystallizes in a spherical form [1-8]. Graphite in this form has a much smaller weakening effect on the matrix than the dispersed graphite flakes, in grey irons [1-8].

Ductile iron is not a single material, but is part of a group of materials that can be produced with a wide range of properties by controlling their microstructure. The common defining characteristic of this group of materials is the shape of graphite, namely nodular (or spheroidal). In nodular graphite irons free graphite is present as spheres or nodules in the as-cast condition [1-4]. The spherical structure of graphite (Figure 1) improves the quality of the cast irons, affecting increased hardness, reliability, supporting significant loads. Nodular irons therefore have considerably higher strength, ductility, and impact values than grey irons [1-8]. Ductile irons determine its properties by ferrite or perlite bases with the presence of nodular graphite inclusions. By close control of melting practice nodular irons can be produced in the as-cast condition over a wide range of section thicknesses with any required matrix structure from fully ferritic to fully pearlitic [1-4].

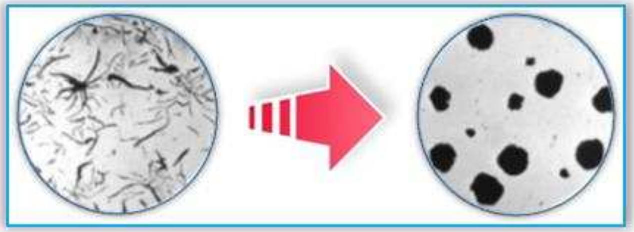


Figure 1
The most favorable form of graphite (minimum effect of loads' concentration)

The formation of nodules is carried out by adding nodulizing elements, Magnesium (0.04-0.12%) most of the time and, less often today, Cerium [1-8]. Magnesium may be added directly to the ladle as Nickel-Magnesium, Nickel-Silicon-Magnesium or Iron-Silicon-Magnesium alloy [1-8].

Phosphorus and Sulphur contents should be as low as possible, as Phosphorus strongly decreases the plasticity and tenacity of the cast irons, and the Sulphur forms compound with the Magnesium (i.e. MgS). Thus contributes to the increase of the nodulizing elements consumption and to the impurification of the cast irons with sulphides [1-8].

Thus, the most important peculiarities of the chemical composition of nodular graphite formations are:

- High Carbon and Silicon content
- Low Phosphorus and Sulphur content
- Proper addition of nodulizing agent (i.e. Magnesium), which ensures the nodularization of graphite

Therefore, quality assurance is limited to good control of the process of elaboration/nodularization of irons [1-8]. It is very important that all elements are in a good correlation, as that ensures the desired quality.

Rolls (Figure 2) – the main and very costly consumables in a rolling mill – are the tools of the rolling trade and the way they are used to execute their duty of deforming steel, in many cases largely determined by the roll's designer [9-16]. They are used in the rolling mill equipment and their performance depend on many factors which include the used materials and the loads to which they are subjected to during technological service [9-16]. The accuracy and speed of working and the roll's life are all related to its peculiar design and choice of materials, which implies a good working knowledge of both the used materials and the loads to which they will be subjected during service [9-16].

Figure 2

The Half-Hard Cast Iron Rolls – As-casting and mechanically processed condition [15] [16]

2 Area of Research

The static-cast nodular iron rolls (Figure 3 and Figure 4) are structurally characterized by the nodular shaped graphite in the microstructure, this feature being essential to the quality and consistency of ductile iron [15] [16]. Ductile rolls are structurally made up of cementite, matrix, and nodular graphite, the free carbon taking the shape of nodules, thereby, eliminating the notch effect of flake graphite and improving the mechanical properties of the cast iron rolls [9-24].

Ductile iron (usually hypereutectic) is produced by ''in-ladle treatment'' with adding Magnesium immediately prior to pouring castings, followed by inoculation in much the same way as for the production of gray cast iron. Cerium and Magnesium additions both produce nodular structures, but the latter has been found to be more adaptable and economical [1-4] [15] [16]. Both elements are de-sulphurizers and nodule formation is not possible until the Sulphur content has been lowered to about 0.02% [5-8] [15] [16].

It is necessary that the Sulphur content be kept <0.01% for successful treatment, because of the affinity of Sulphur for Magnesium (forming Mg2S), thus removing elemental Magnesium from the melt [5-8]. Finally, the nodular irons are inoculated with 0.4-0.8% Ferro-Silicon after nodulizing to refine the structure and minimize chilling [5-8] [15] [16].

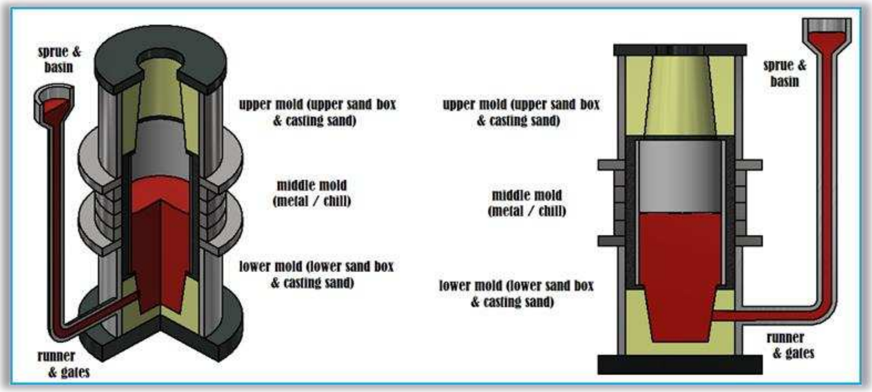

sprue & basin
upper mold (upper sand box & casting sand)
middle mold (metal / chill)
lower mold (lower sand box & casting sand)
runner & gates
upper mold (upper sand box & casting sand)
middle mold (metal / chill)
lower mold (lower sand box & casting sand)
sprue & basin
runner & gates

Figure 3
The Half-Hard Cast Iron Rolls – The static-cast process [15] [16]

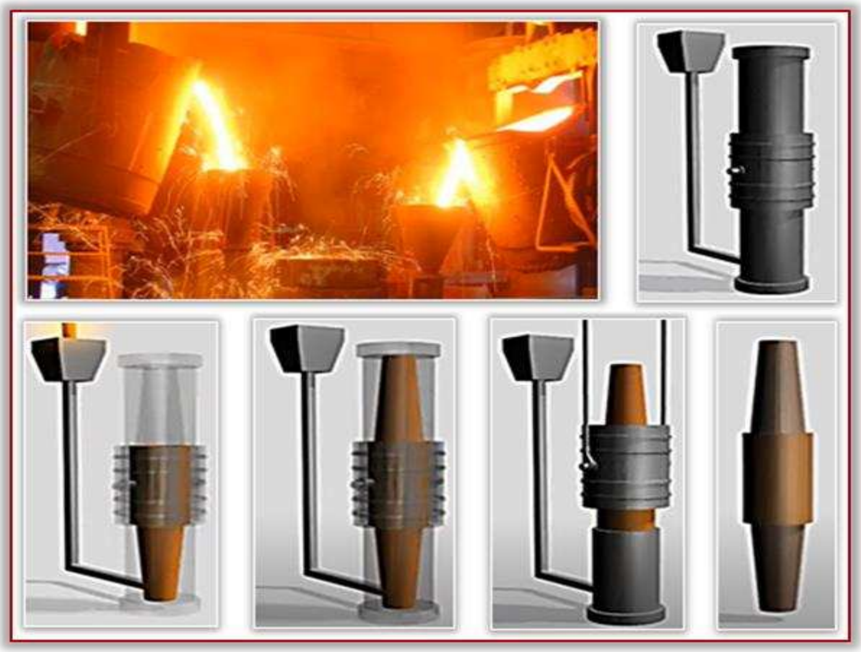


Figure 4
The Half-Hard Cast Iron Rolls – The casting & operational phases [15] [16]

The recommendations for the increase of the duration of exploitation and remove of the damages through the accidental rupture of rolls from the stands of lamination, the attenuation of rolls thermal fatigue, the avoiding of thermal shocks caused damages and their rational exploitation are actuality issues that must be continuously researched [9-24]. In this trend is situated the research of the cast-iron rolls (Figure 5) [15-24]. The quality of rolls is determined through hardness and through wear resistance, last index having a special importance for all modern rolling mills [9-24]. The presents of graphite in working surface (body of rolls) assures the friction coefficient necessary to obtain quality laminate [15-24].

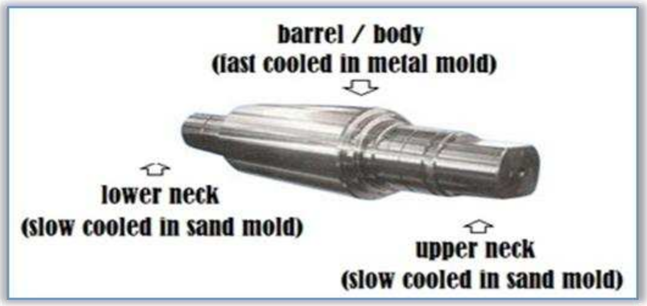

barrel / body
(fast cooled in metal mold)
lower neck
(slow cooled in sand mold)
upper neck
(slow cooled in sand mold)

Figure 5
The Half-Hard Cast Iron Rolls [15] [16]

The rolls must present high hardness at the body of rolls and lower in the core and the neck's, adequate with mechanical resistance and in the high work temperature. If in the body, the hardness is guaranteed by the quantities of cementite in the structure of irons, the core of rolls must content graphite, to assure this property [9-24]. The research includes half–hard cast rolls, from nodular graphite irons, with the half-hard body of 40-150 mm depth [15] [16] [25]. The macrostructure is not imposed (except for the nodular graphite irons, where a spherical shape of the graphite is required), conditioned by the adequate quantities of cementite in the body and graphite in the core and on the necks [9-24]. Thus, the optimal additions in these elements can be determined to assure the proper hardness (Table 1 and Table 2).

Table 1
The Recommended Chemical Composition of the Half-hard Cast Iron Rolls [25]

Chemical Composition, (%)								
Carbon [C]	Silicon [Si]	Manganese [Mn]	Phosphorus [P]	Sulphur [S]	Nickel [Ni]	Chromium [Cr]	Molybdenum [Mo]	Magnesium [Mg]
3.0-3.5	1.2-2.5	0.1-0.7	max 0.15	max 0.02	1.5-2.5	max 0.8	0.3-0.5	0.02-0.04

This study analyses iron rolls cast in the static procedure, in combined forms (iron chill, for the barrel and molding sand, for the necks of the rolls) [15-24]. The research included 108 rolls from the half-hard class – 0 and 1 hardness class, 219-347 Brinell units on body (working surface) and 195-271 Brinell units in core and on the necks of rolls (Table 3) [15] [16] [25].

Table 2

The Recommended Hardness of the Half-hard Cast Iron Rolls [25]

Class of Hardness	Recommended Hardness for these Rolls [Brinell Hardness]	
	on Body (Rolling Surface) of Rolls	in Core and on the Neck's of Rolls
0	218-286	195-271
1	294-347	195-271

Table 3

The Chemical Composition and the measured Hardness of the Half-hard Cast Iron Rolls

Chemical Composition, (%)			
Carbon [C]	3.22-3.42	Nickel [Ni]	1.49-2.22
Silicon [Si]	1.72-2.19	Chromium [Cr]	0.36-0.72
Manganese [Mn]	0.62-0.79	Molybdenum [Mo]	0.18-0.28
Phosphorus [P]	0.130-0.165	Magnesium [Mg]	0.021-0.029
Sulphur [S]	0.011-0.024		
Hardness, [Brinell units]			
on the Necks		219-276	
on the Body		282-352	

3 Research Methodology

The mathematical modeling was applied, taking into consideration the industrial data obtained from the rolls industry, as well as, the cast iron roll's requirements [15] [16]. Therefore, it is suggested to use a mathematical interpretation on influence of the ductile iron's particulate elements (Phosphorus, Sulphur and Magnesium) over the hardness on the rolls rolling surface (barrel or body) and on the necks [15-24].

The performed research had in view to obtain correlations between the Half-Hard Cast Iron Rolls' Hardness and the nodular iron's permanent elements, defined by two of the main elements which have a major influence on the microstructure (i.e. Sulphur and Phosphorus), respectively by the nodulizing element (i.e. Magnesium). The performed research is structured in two specific cases, based on the different values measured on the roll's components (the two necks and the body/rolling surface) [15-24].

The Half-Hard Cast Iron Rolls' chemical composition and their hardness variation limits (measured on necks and on the body) are presented in Table 4. Also, the average values and the deviations of variables are presented in Table 5.

Table 4

The Half-hard cast iron rolls' chemical composition and the hardness variation limits ([HB]$_{(necks)}$ = f([S],[P],[Mg]) and [HB]$_{(body)}$ = f([S],[P],[Mg])

Case 1: Modelling on data measured on the roll's necks							
Sulphur [S]		Phosphorus [P]		Magnesium [Mg]		Hardness [HB]$_{(necks)}$	
[S]$_{inf}$	[S]$_{sup}$	[P]$_{inf}$	[P]$_{sup}$	[Mg]$_{inf}$	[Mg]$_{sup}$	[HB]$_{inf}$	[HB]$_{inf}$
0.011	0.024	0.128	0.165	0.021	0.031	219	276
Case 2: Modelling on data measured on the roll's body (rolling surface)							
Sulphur [S]		Phosphorus [P]		Magnesium [Mg]		Hardness [HB]$_{(body)}$	
[S]$_{inf}$	[S]$_{sup}$	[P]$_{inf}$	[P]$_{sup}$	[Mg]$_{inf}$	[Mg]$_{sup}$	[HB]$_{inf}$	[HB]$_{inf}$
0.011	0.024	0.128	0.165	0.021	0.031	280	352

Table 5

The Half-hard cast iron rolls' chemical composition and the hardness average values and deviations of variables ([HB]$_{(necks)}$ = f([S],[P],[Mg]) and [HB]$_{(body)}$ = f([S],[P],[Mg])

Case 1: Modelling on data measured on the roll's necks							
Sulphur [S]		Phosphorus [P]		Magnesium [Mg]		Hardness [HB]$_{(necks)}$	
[S]$_{med}$	deviation	[P]$_{med}$	deviation	[Mg]$_{med}$	deviation	[HB]$_{(necks)med}$	deviation
0.0179	0.0036	0.1515	0.0107	0.0256	0.0029	251.52	13.622
Case 2: Modelling on data measured on the roll's body (rolling surface)							
Sulphur [S]		Phosphorus [P]		Magnesium [Mg]		Hardness [HB]$_{(body)}$	
[S]$_{med}$	deviation	[P]$_{med}$	deviation	[Mg]$_{med}$	deviation	[HB]$_{(body)med}$	deviation
0.0179	0.0036	0.1515	0.0107	0.0256	0.0029	308.32	22.107

A rigorous foundation of the existence of a correlation and the existence of a model describing the correlation between variables, also called a regression model, can be made on the basis of the calculation and interpretation of statistical indicators [15-24]. Several steps will be taken, such as:

- Check the existence of a correlation
- Establish the mathematical shape of the model
- Analysis of the empirical regression curve
- Determine the parameters that appear in the model equation
- Use the model to forecast calculations

There are also parameters that measure the correlation (degree of association) between two qualitative variables, parameters based on the occurrence frequencies of variable values and not on values [15] [16]. Once the existence of the correlation between variables is established, we can proceed to the establishment of the regression model describing the correlation [15-24].

The mathematical modeling was applied based on the differentiation between the component parts of the rolls [15] [16] [25]. Starting from the rolling rolls' aspect, the shape of the rolls, the areas of technological interest and the structure that ensures the operational mechanical properties, it has been developed, by modeling, the mathematical description of direct influences and finally, by successive determinations, an optimum chemical composition which assures the desired rolls' hardness [15-24]. The multiple regression was used to predict the value of a variable based on the value of two or more other variables [15-24]. The study of a regression model involves the following aspects:

- Determination of regression hyper-surface, specific to multiple regression with several variables
- Determination of correlation coefficients
- Determination of deviation from the regression surface
- Determining the coordinates of the optimal points, for which there are desired values
- Determination of regression areas, specific to regression with 2 variables
- The graphic representation of the regression curve based on observed data
- Verification on the correlation graphs of the optimal range

In our statistical modelling, the regression analysis was used for estimating the relationships among the proper hardness (Hardness [HB]) and these elements (Phosphorus, Sulphur and Magnesium). To determine to what extent independent variables contribute to the modification of the dependent variable a multiple regression model was developed and will determine whether it can be considered valid, i.e. whether or not there is a correlation between a mechanical property (rolls' hardness), the level of concentrations of permanent chemical elements (Phosphorus and Sulphur) and the concentrations of elements added to the nodulizing treatment with Magnesium, characterized by certain values of independent variables.

4 Results of the Statistical Modeling

Statistical modeling determines the coordinates of the optimal point ([S],[P],[Mg]), for which $[HB]_{(necks)}$ and $[HB]_{(body)}$ has desired values (Table 6).

Table 6

The coordinates of the optimal point

Case 1: Modelling on data measured on the roll’s necks			
$[S]_{statistical}$	$[P]_{statistical}$	$[Mg]_{statistical}$	$[HB]_{(necks)statistical}$
0.0191	0.1472	0.0259	260.9375
Case 2: Modelling on data measured on the roll’s body (rolling surface)			
$[S]_{statistical}$	$[P]_{statistical}$	$[Mg]_{statistical}$	$[HB]_{(body)\ statistical}$
0.0212	0.1562	0.0235	327.1099

The correlations between the hardness of the Half-hard cast iron rolls and the defined three chemical elements are studied ($[HB]_{(necks)}$ = f([S],[P],[Mg]) and $[HB]_{(body)}$ = f([S],[P],[Mg]), using the values presented in Table 4. Two polynomial type of correlation was revealed, presented in the equation (1) and (2).

The proper mathematical correlation, in the case of $[HB]_{(necks)}$ = f([S],[P],[Mg]), is given by the equation of regression hyper-surface (1), where the correlation coefficient is rf = 0.6816 and the deviation from the regression surface is sf = 9.9675. The proper mathematical correlation, in the case of $[HB]_{(body)}$ = f([S],[P],[Mg]), is given by the equation of regression hyper-surface (2), where the correlation coefficient is rf = 0.6841 and the deviation from the regression surface is sf = 16.1240.

$$[HB]_{(necks)} = -1729.1599\,[S]^2 - 33258.1844\,[P]^2 - 14187.7104\,[Mg]^2 - 72143.1146\,[S][P] - 34036.2106\,[P]\,[Mg] - 10818.3003\,[Mg][S] + 45269.1696\,[S] + 12044.2871\,[P] + 33001.3626\,[Mg] - 1484.7255 \tag{1}$$

$$[HB]_{(body)} = -5523.0405\,[S]^2 - 44607.2162\,[P]^2 - 6146.3995\,[Mg]^2 + 14432.6465[S][P] - 41587.5438\,[P][Mg] - 13741.9865\,[Mg][S] + 53316.7961\,[S] + 14603.5068\,[P] + 64322.9104\,[Mg] - 2130.4001 \tag{2}$$

Since this hyper-surface cannot be represented in the 4th dimensional space, I resorted to replacing, successively, an independent variable [15] [16]. In order to determine the limits of graphic representation, it is used to replace, successively, a variable independent with the values of the limits of the chemical composition, presented above in Table 4. These surfaces in the 3rd dimensional space are governed by the eq. (1.1) – (1.6), respectively equations (2.1) – (2.6).

$$[HB]_{(necks)}[S]_{inf} = -33258.1844\,[P]^2 - 14187.7104\,[Mg]^2 - 34036.2106\,[P][Mg] + 10625.1483\,[P] + 11721.3703\,[Mg] - 661.1456 \tag{1.1}$$

$$[HB]_{(necks)}[S]_{sup} = -33258.1844\,[P]^2 - 14187.7104\,[Mg]^2 - 34036.2106\,[P][Mg] + 10885.0937\,[P] + 15619.2532\,[Mg] - 801.9896 \tag{1.2}$$

$$[HB]_{(necks)}[P]_{inf} = -\ 14187.7104\ [Mg]^2 - 1729.1599\ [S]^2 - 10818.3003\ [Mg][S] + 27663.5758\ [Mg] + 33955.2026\ [S] - 413.8319 \qquad (1.3)$$

$$[HB]_{(necks)}[P]_{sup} = -\ 14187.7104\ [Mg]^2 - 1729.1599\ [S]^2 - 10818.3003\ [Mg][S] + 28027.6574\ [Mg] + 34726.9095\ [S] - 434.8887 \qquad (1.4)$$

$$[HB]_{(necks)}[Mg]_{inf} = -\ 1729.1599[\ S]^2 - 33258.1844\ [P]^2 - 72143.1146\ [S][P] + 16011.4863\ [S] + 11123.7535\ [P] - 695.9543 \qquad (1.5)$$

$$[HB]_{(necks)}[Mg]_{sup} = -\ 1729.1599\ [S]^2 - 33258.1844\ [P]^2 - 72143.1146\ [S][P] + 19214.6491\ [S] + 11224.5345\ [P] - 772.1924 \qquad (1.6)$$

$$[HB]_{(body)}\ [S]_{med} = -\ 44607.2162\ [P]^2 - 6146.3995\ [Mg]^2 - 41587.5438\ [P][Mg] + 14887.4137\ [P] + 37291.7653\ [Mg] - 1295.3177 \qquad (2.1)$$

$$[HB]_{(body)}[S]_{med} = -\ 44607.2162\ [P]^2 - 6146.3995[Mg]^2 - 41587.5438\ [P][Mg] + 14835.4102\ [P] + 42243.0941[Mg] - 1416.3042 \qquad (2.2)$$

$$HB]_{(body)}[P]_{med} = -\ 6146.3995\ [Mg]^2 - 5523.0405\ [S]^2 - 13741.9865\ [Mg][S] + 57800.8729\ [Mg] + 55580.2205\ [S] - 937.2776 \qquad (2.3)$$

$$[HB]_{(body)}[P]_{med} = -\ 6146.3995\ [Mg]^2 - 5523.0405\ [S]^2 - 13741.9865\ [Mg][S] + 58245.7302[Mg] + 55425.8361[S] - 948.9314 \qquad (2.4)$$

$$[HB]_{(body)}[Mg]_{med} = -\ 5523.0405\ [S]^2 - 44607.2162\ [P]^2 + 14432.6465\ [S][P] + 16151.9009\ [S] + 13478.7422\ [P] - 840.3223 \qquad (2.5)$$

$$[HB]_{(body)}[Mg]_{med} = -\ 5523.0405\ [S]^2 - 44607.2162\ [P]^2 + 14432.6465\ [S][P] + 20220.7538\ [S] + 13601.8825\ [P] - 937.7295 \qquad (2.6)$$

In the equation of hyper-surfaces (1) and (2), it is used to replace, successively, a variable independent with its mean value. These surfaces, which belong to the whole space with 3rd dimensions, can be represented and interpreted by technologists. Therefore, the independent variables were successively replaced with their average values (i.e. [S]med, [P]med and [Mg]med, Table 5).

A polynomial type of correlations was revealed, which have the following general forms, presented in the equations (1.7) – (1.9), respectively (2.7) – (2.9).

$$[HB]_{(necks)}[S]_{med} = -\ 33258.1844\ [P]^2 - 14187.7104\ [Mg]^2 - 34036.2106\ [P][Mg] + 10755.1212\ [P] + 13670.3117\ [Mg] - 731.0063 \qquad (1.7)$$

$$[HB]_{(necks)}[P]_{med} = -\ 14187.7104\ [Mg]^2 - 1729.1599\ [S]^2 - 10818.3003\ [Mg][S] + 27845.6166\ [Mg] + 34341.0561\ [S] - 423.4089 \qquad (1.8)$$

$$[HB]_{(necks)}[Mg]_{med} = -\ 1729.1599\ [S]^2 - 33258.1844\ [P]^2 - 72143.1146\ [S][P] + 17613.0677\ [S] + 11174.1442\ [P] - 733.7624 \qquad (1.9)$$

$$[HB]_{(body)}[S]_{med} = -\ 44607.2162\ [P]^2 - 6146.3995\ [Mg]^2 - 41587.5438\ [P][Mg] + 14861.4122\ [P] + 39767.4297\ [Mg] - 1354.0182 \qquad (2.7)$$

$$[HB]_{(body)}[P]_{med} = -\,6146.3995\,[Mg]^2 - 5523.0405\,[S]^2 - 13741.9865\,[Mg][S] + 58023.3016\,[Mg] + 55503.0283\,[S] - 941.8285 \qquad (2.8)$$

$$[HB]_{(body)}[Mg]_{med} = -\,5523.0405\,[S]^2 - 44607.2162\,[P]^2 + 14432.6465\,[S][P] + 18186.3274\,[S] + 13540.3122\,[P] - 887.6787 \qquad (2.9)$$

5 Graphical Addenda

The 3th dimensional regression surfaces, described by the governing equations (1.1) – (1.6) and (2.1) – (2.6), respectively, are represented graphically in Figure 6, case of $[HB]_{(necks)} = f([S],[P],[Mg]))$ and Figure 7, case of $[HB]_{(body)} = f([S],[P],[Mg]))$.

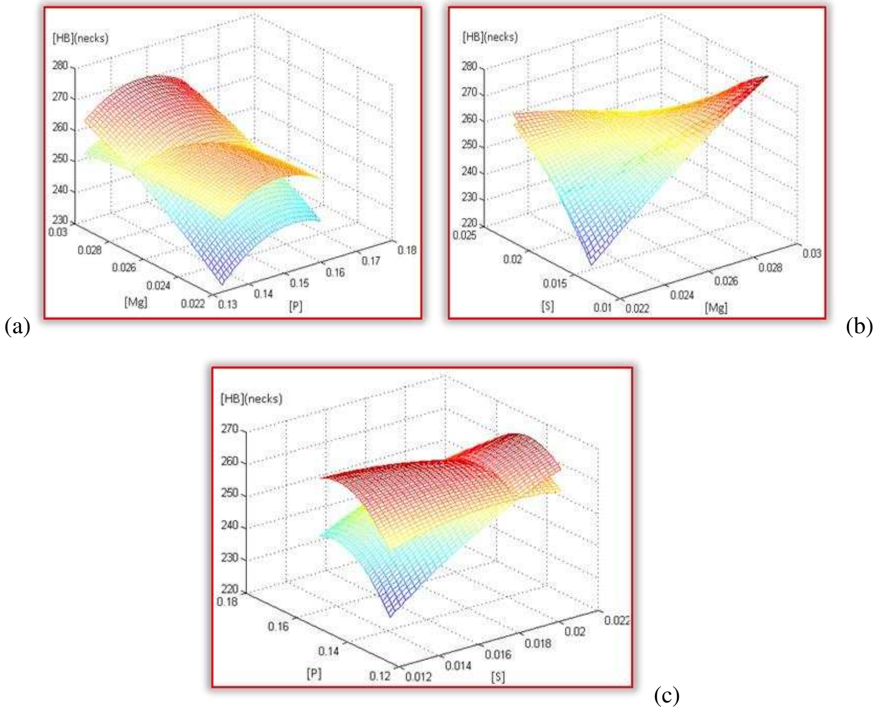


Figure 6

Regression surfaces in case of $[HB]_{(necks)} = f([S],[P],[Mg])$, according to the equations (1.1)–(1.6). (a) the regression surface described by the industrial data, when [S] =[S]inf and [S]=[S]sup; (b) the regression surface described by the industrial data, when [P] =[P]inf and [P]=[P]sup; (c) the regression surface described by the industrial data, when [Mg] =[Mg]inf and [Mg]=[Mg]sup

The regression surfaces, described by the eq. (1.1) – (1.9), respectively eq. (2.1) – (2.9), are presented in Figures 8-10 and Figures 11-13. The correlation charts and the contour lines which define the requested limits of the proper hardness, in cases of $[HB]_{(necks)} = f([S],[P],[Mg])$ and $[HB]_{(necks)} = f([S],[P],[Mg])$ are presented in the Figures 14-19.

In this sense, the rolls' hardness variations on the necks and core, described by these elements, are presented in Figures 14-16, determined by Matlab, using the polynomial equations presented in the eq. (1.7), (1.8) and (1.9). The rolls' hardness variations on the body, are presented in Figures 17-19, using the polynomial equations presented in the equations (2.7), (2.8) and (2.9).

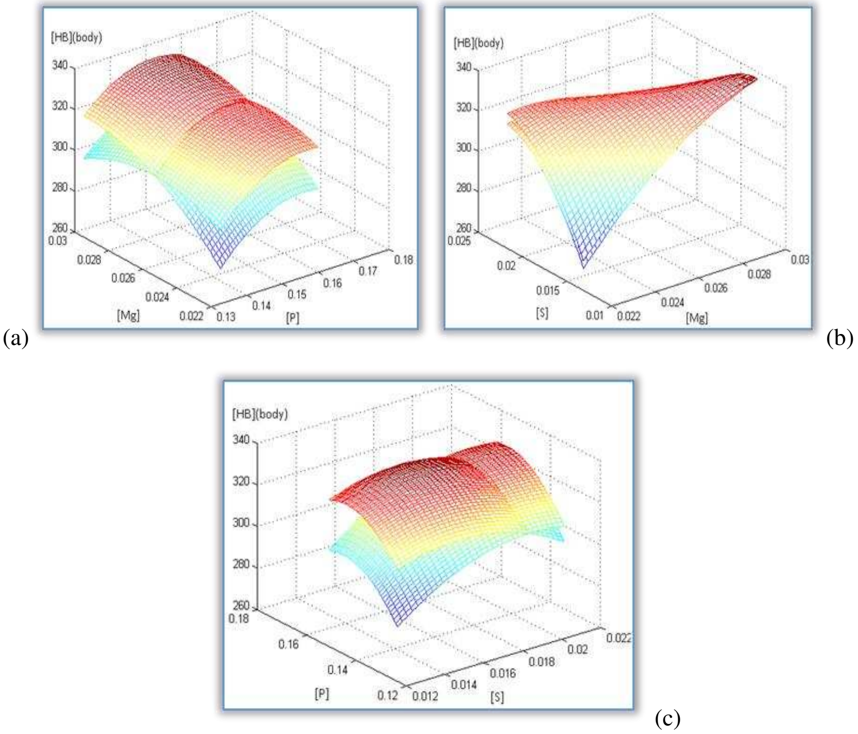


Figure 7

Regression surfaces in case of $[HB]_{(body)} = f([S],[P],[Mg])$, according to the equations (2.1)–(2.6)

(a) the regression surface described by the industrial data, when [S] =[S]inf and [S]=[S]sup;

(b) the regression surface described by the industrial data, when [P] =[P]inf and [P]=[P]sup;

(c) the regression surface described by the industrial data, when [Mg] =[Mg]inf and [Mg]=[Mg]sup

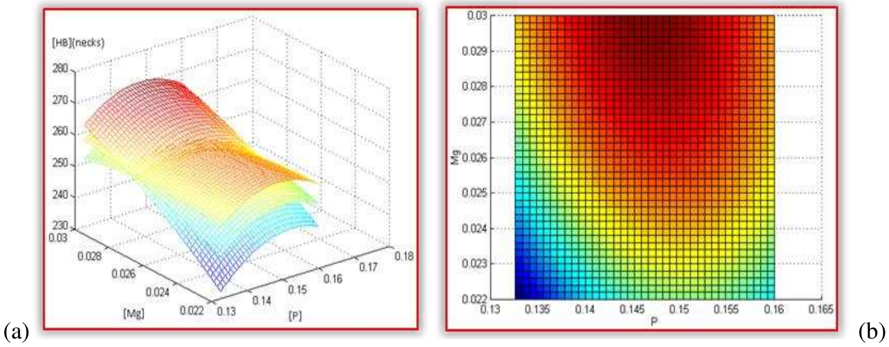


Figure 8

Regression surfaces in case of $[HB]_{(body)} = f([S],[P],[Mg])$, according to the equation (1.7)

(a) the regression surfaces; (b) color mapping of the technological areas

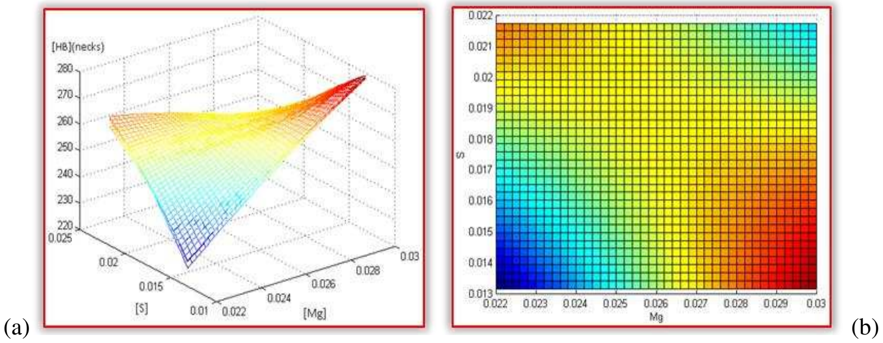


Figure 9

Regression surfaces in case of $[HB]_{(body)} = f([S],[P],[Mg])$, according to the equation (1.8)

(a) the regression surfaces; (b) color mapping of the technological areas

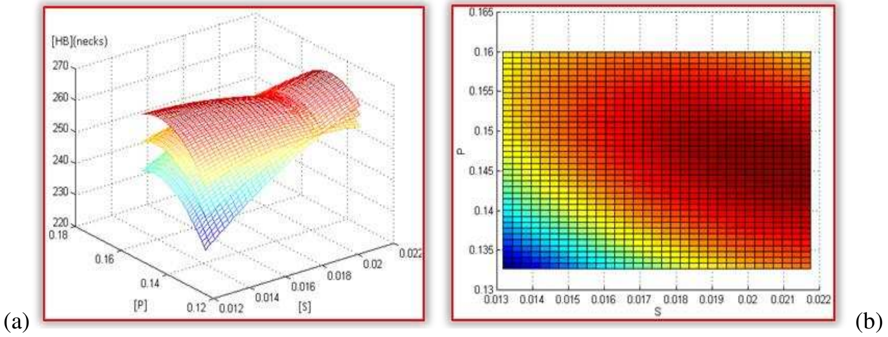


Figure 10

Regression surfaces in case of $[HB]_{(body)} = f([S],[P],[Mg])$, according to the equation (1.7)

(a) the regression surfaces; (b) color mapping of the technological areas

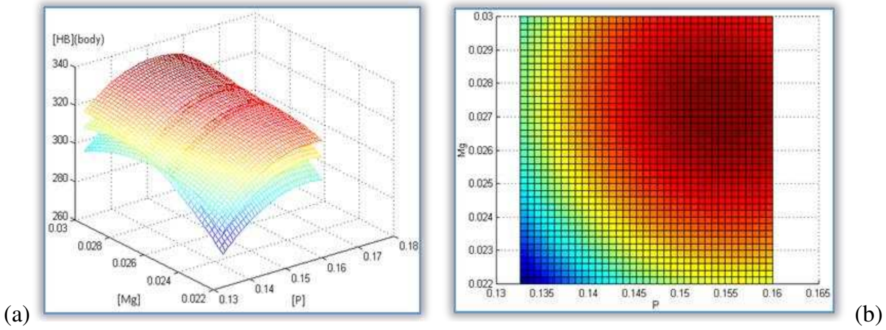


Figure 11

Regression surfaces in case of $[HB]_{(necks)} = f([S],[P],[Mg])$, according to the equation (2.7)

(a) the regression surfaces; (b) color mapping of the technological areas

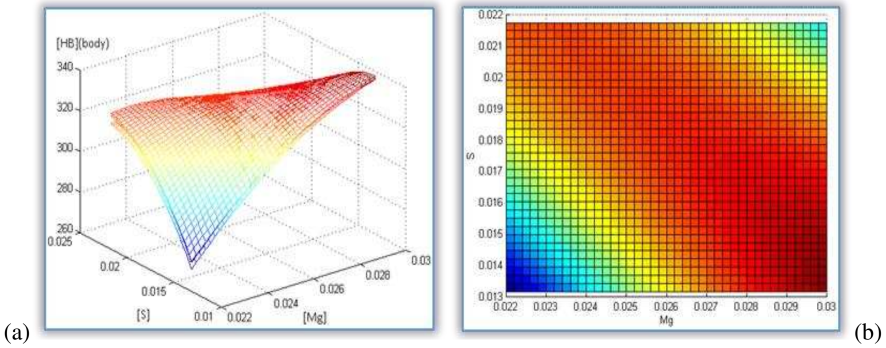


Figure 12

Regression surfaces in case of $[HB]_{(necks)} = f([S],[P],[Mg])$, according to the equation (2.8)

(a) the regression surfaces; (b) color mapping of the technological areas

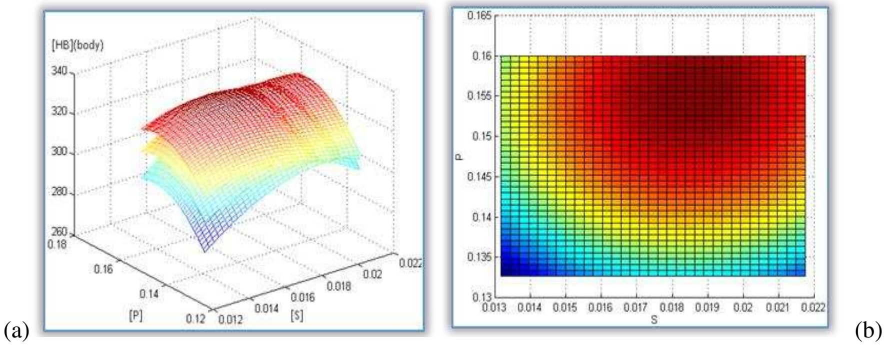


Figure 13

Regression surfaces in case of $[HB]_{(necks)} = f([S],[P],[Mg])$, according to the equation (2.9)

(a) the regression surfaces; (b) color mapping of the technological areas

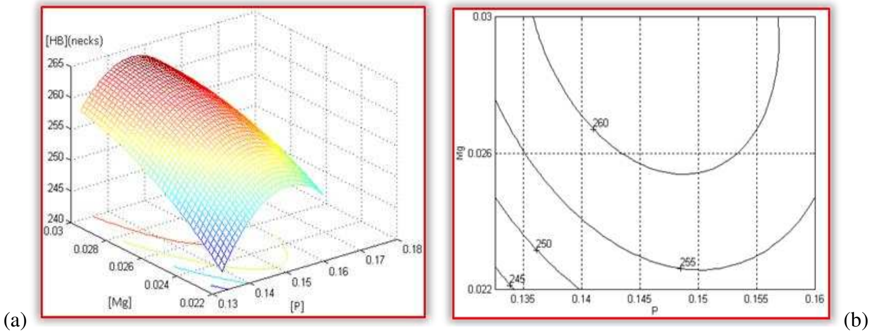


Figure 14

Correlation charts in case of $[HB]_{(necks)} = f([S],[P],[Mg])$, when [S]=[S]med. (a) regression surface for $[HB]_{(necks)} = [HB]_{(necks)}([S],[P],[Mg]_{med})$; (b) contour lines for $[HB]_{(necks)} = [HB]_{(necks)}([S]_{med},[P],[Mg])$

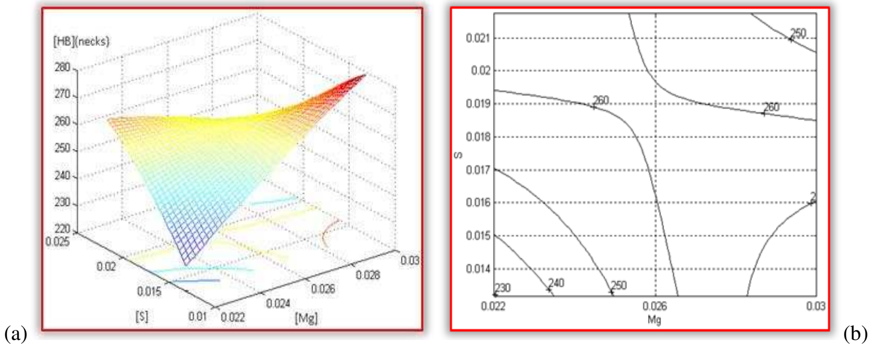


Figure 15

Correlation charts in case of $[HB]_{(necks)} = f([S],[P],[Mg])$, when [P]=[P]med. (a) regression surface for $[HB]_{(necks)} = [HB]_{(necks)}([S],[P],[Mg]_{med})$ (b) contour lines for $[HB]_{(necks)} = [HB]_{(necks)}([S],[P]_{med},[Mg])$

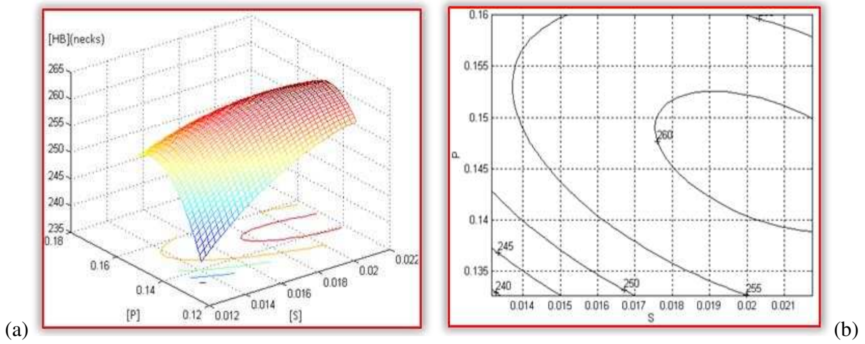


Figure 16

Correlation charts in case of $[HB]_{(necks)} = f([S],[P],[Mg])$, when [Mg]=[Mg]med. (a) regression surface for $[HB]_{(necks)} = [HB]_{(necks)}([S],[P],[Mg]_{med})$; (b) contour lines for $[HB]_{(necks)} = [HB]_{(necks)}([S],[P],[Mg]_{med})$

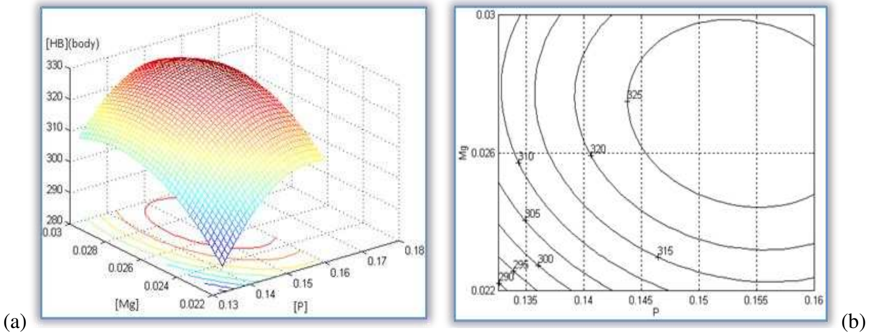


Figure 17

Correlation charts in case of $[HB]_{(necks)}$ = f ([S],[P],[Mg]), when[S]=[S]med. (a) regression surface for $[HB]_{(necks)}$ = $[HB]_{(necks)}$([S],[P],$[Mg]_{med}$); (b) contour lines for $[HB]_{(necks)}$ = $[HB]_{(necks)}$($[S]_{med}$,[P],[Mg])

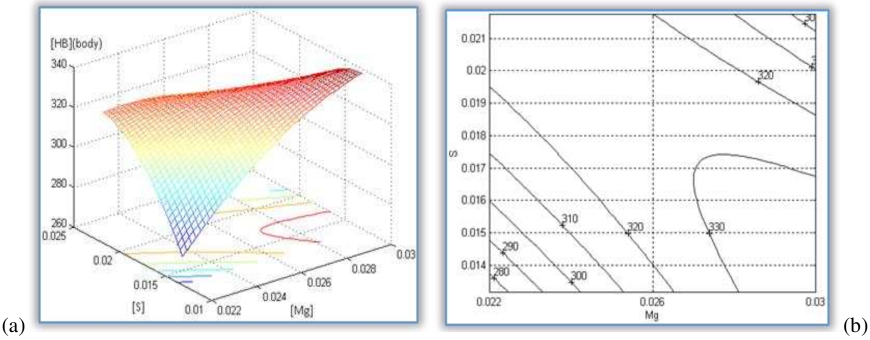


Figure 18

Correlation charts in case of $[HB]_{(necks)}$ = f ([S],[P],[Mg]), when [P]=[P]med. (a) regression surface for $[HB]_{(necks)}$ = $[HB]_{(necks)}$([S],[P],$[Mg]_{med}$); (b) contour lines for $[HB]_{(necks)}$ = $[HB]_{(necks)}$([S],$[P]_{med}$,[Mg])

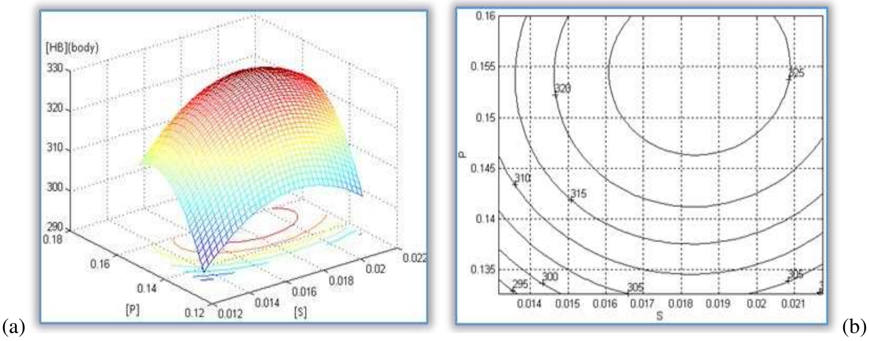


Figure 19

Correlation charts in case of $[HB]_{(necks)}$ = f ([S],[P],[Mg]), when [Mg]=[Mg]med. (a) regression surface for $[HB]_{(necks)}$ =$[HB]_{(necks)}$([S],[P],$[Mg]_{med}$); (b) contour lines for $[HB]_{(necks)}$ = $[HB]_{(necks)}$([S],[P],$[Mg]_{med}$)

6 Discussion

Regarding the multiple regression analysis, in order to understand the relationships between the variables and their relevance to the problem being studied, are the following remarks:

- The problem of regression starts from the existence of a data set on two or more random variables, the purpose of modeling being the description of the relationship between them, i.e. determining function f, in order to forecast the values of the dependent variable in relation to the values of the explanatory variables. This problem arises only when there is a real link between the variables, based on the nature of the underlying phenomena.
- The coefficient shows the role played by all exogenous (dependent) variables on the evolution of the endogenous variable (independent). The correlation coefficient is interpreted as follows: a high value, close to 1, indicating a good adjustment of the data, while a value close to 0, indicates a weak adjustment. In both cases ($[HB]_{(necks)}$ = f ([S],[P],[Mg])) and $[HB]_{(body)}$ = f ([S],[P],[Mg])), the values of the correlation coefficient (rf = 0.6816, respectively rf = 0.6841) indicate a fairly good correlation between the studied variables.
- The used model of regression belongs to the category of stochastic (statistical) models, in which all the explanatory factors of a phenomenon, which do not find their place directly in the model, appear accumulated in the form of a random variable called error. A variable (an output parameter) that quantifies the studied phenomenon can be explained by regression on one or more explanatory factors (input parameters). All explanatory factors that are not sufficiently relevant to the output parameter enter into the model in the cumulative form of the error. In this model, the residues (errors) do not behave randomly and are quite large.
- The technological domain area of proper hardness is presented in Figures 14-16, respectively in Figures 17-19. The relationships that determine the technological areas are useful because they can indicate a predictive relationship that can be exploited in practice.

In summary, the recent study of a regression model ($[HB]_{(necks)}$ = f ([S],[P],[Mg]) and $[HB]_{(body)}$ = f ([S],[P],[Mg]) involves the following aspects:

- Establish the analytical limits, defined statistically
- Determining the mean value and deviations of the variables from the mean values
- Determination the regression hyper-surface’s equations, correlation coefficients and regression surface deviations, resulting the eq. (1) and (2)
- Establishing the optimum value coordinates, for which the hardness has desired (optimal) values

- Establishing the regression surface's equations, on the path of successive imposition of limit values and value chemical composition , resulting the eq. (1.1-1.9) and (2.1-2.9)
- Representation the regression surfaces, presented as rich graphical addenda, in section 5 (Figures 6-13)
- Determination the correlation charts and the level curves (contour lines) and mapping the technological areas, presented in Figures 14-16 and in Figures 17-19
- Validate the statistically determined optimal domain in the correlation charts, in which it must be found, and in our case are found, the values determined statistically and presented in Table 6.

Having in view the complex melting process of the ductile irons destined to the half-hard cast iron rolls manufacturing, followed by a proper nodulizing process of these irons in the ladle treatment, I have the following technological remarks:

- As presented in the previous works [15-24], the chemical composition is a very important factor in the assurance of mechanical properties, especially the hardness, of these important iron products (i.e. rolls), parts of the rolling equipment. In this sense, it was declared constantly, that one of the basic factors that determine the rolls' structure is the chemical composition. Between these elements, these research show that the permanent elements like Phosphorus and Sulphur and the nodulizing agent content like Magnesium have an important technological role in assurance of these exploitation properties.
- Therefore, the desired properties of the rolls can be assured besides a proper content in Phosphorus and Sulphur, in the charging and melting process, respectively in Magnesium, in the ladle process stage, having in view the behaviour of these elements. In this sense, a proper addition of the Phosphorus and Sulphur contents are provided by the basic metallic charges used in the melting process, having in view the technological prescriptions in low Phosphorus and Sulphur content of the ductile irons. Also, important is the assuring the graphite spheres or nodules (in quantity and as form) in the as-cast condition, by adding Magnesium, which will improve the quality of these cast irons, offering an increased hardness of the rolls.

Concluding Remarks

The improved properties obtained with high-purity raw materials in making ductile iron should stimulate further investigations particularly related to obtaining the desired properties of cast products, such the rolling rolls. The mechanical properties of iron are not only determined by composition but also greatly influenced by foundry practice, particularly cooling rate in the casting. With the exception of Magnesium or other nodularizing elements in nodular iron, it is possible through variations in melting and foundry practice to produce different properties from the same composition. These properties are produced by increasing the alloy content mainly Nickel, Molybdenumand Chromiummodifying the matrix structure [15-24]. But, with a high percentage of graphite nodules present in the structure, the mechanical properties are determined by the ductile iron matrix. Hardness values, usually offered as additional information, and impact properties, specified only for certain grades, complete most specifications.

Based on the results obtained in the performed statistical research, it is concluded that prediction of the exploitation properties of rolls, based on the melting process and the additional "in-ladle" treatment, is a prerequisite for the cast iron roll's manufacturing. The statistical modelling by multivariate regression analysis can be used successfully to optimize the chemical composition of irons destined to rolls manufacturing. In this way, this method is very helpful to predict the cast roll's performance.

Therefore, the chemical composition of cast iron is a key-factor that largely determines mechanical properties the resulting rolls. By close control of analysis of the melting process and the additional nodulizing and inoculation practices, nodular irons can be produced in the as-cast condition over a wide range of section thicknesses with any required matrix structure from fully ferritic to fully pearlitic.

These rolls are produced in wide range of composition to satisfy the rolling mills requirement, nodular irons being much less section sensitive than grey irons. Phosphorus and Sulphur content should be as low as possible. Sulphur forms with Magnesium compounds and thus contributes to the increase of the consumption of nodulizing agent, using it inappropriately and unnecessarily. An increase in the concentration of Phosphorus in the composition causes the appearance of cracks when the composition is cooled. In addition, this element causes damage to other mechanical properties. Also, many properties are influenced by the mechanisms of primary and secondary crystallization. Ductile irons determine their properties by ferritic or pearlitic bases with the presence of nodular graphite inclusions, most of them being primary or secondary products of treatment with Magnesium.

In the complete understanding of technological reality it is often necessary to know and understand the existing correlations between two or more phenomena, quantified by different variables. For example, in order to apply a correct quality strategy on milling rolls, it is necessary to know whether there is a correlation between the acquired properties of the rolls and the main factors which influenced their manufacture.

Acknowledgement

The research includes own studies, analyses and research, regarding the quality improvement by modeling of the statistical data from the half-hard cast iron rolls used in the steel rolling. At the fundament of this statistical research upon the cast iron rolls are several relevant scientific studies [15] [16] and approaches [17-24] in which the author is involved in the last 25 years.

References

[1] N. S. Tiedje, Solidification, processing and properties of ductile cast iron, Materials Science and Technology, 26/5 (2010), pp. 505-514

[2] J. Lacaze, Trace elements and graphite shape degeneracy in nodular graphite cast irons, International Journal of Metalcasting, 11/1 (2017), pp 44-51

[3] T. Skaland, Production of ductile cast iron from high purity charge materials, Scandinavian Journal of Metallurgy, 21/2 (1992), pp. 63-67

[4] I. Ripoşan, M. Chişamera, S. Stan, G. Grasmo, C. Hartung, D. White, Iron quality control during melting in coreless induction furance, AFS Transactions, 117 (2009), pp. 423-434

[5] O. M. Suarez, R. D. Kendrick, C. R. Loper, A study of sulphur effect in high silicon ductile irons, International Journal of Cast Metals Research, 13/3 (2000), pp. 135-145

[6] M. Chisamera, I. Riposan, M. Barstow, Sulphur inoculation of Mg–treated cast iron – An efficient possibility to obtain compacted graphite cast iron and to improve graphite nucleation ability in ductile iron, AFS Transactions, 104 (1996) pp. 581-588

[7] A. Oyetunji, S. O. Omole, Achievement of nodules in ductile iron having sulphur content not less than 0.07% weight, Annals of the Faculty of Engineering Hunedoara, 12/4 (2014) 42-46

[8] K. Janerka, J. Jezierski, D. Bartocha, J. Szajnar, Analysis of ductile iron production on steel scrap base, International Journal of Cast Metals Research, 27/4 (2014) pp. 230-234

[9] W. H. Betts, H. L. Baxter, Rolls used in today's rolling mills, Iron and Steelmaker, 18/2 (1991) pp. 42-43

[10] S. Spuzic, K. N. Strafford, C. Subramanian, G. Savage, Wear of hot rolling mill rolls: an overview, Wear, 176/2 (1994) pp. 261-271

[11] Z. Stradomski, A. Pirek, S. Stachura, Studying possibilities to improve the functional properties of metallurgical rolls, 8/1 (2008) pp. 313-316

[12] E. Kerr, R. Webber, D. McCaw, Roll performance-technical overview and future outlook, Ironmaking & steelmaking, 31/4 (2004) pp. 295-299

[13] J. Krawczyk, J. Pacyna, Influence of a matrix on properties of mottled cast iron applied for mill rolls, Archives of Foundry Engineering, 10/3 (2010) pp. 45-50

[14] J. Krawczyk, Structural causes of defects in a cast iron mill roll, Archives of Foundry Engineering, 8/2 (2008) pp. 93-98

[15] I. Kiss, The quality of rolling mills rolls cast by irin with nodular graphite, Mirton Publishing House, Timisoara, 2005

[16] I. Kiss, Rolling rolls – Approaches of quality in the multidisciplinary research, Mirton Publishing House, Timisoara, 2008

[17] I. Kiss, St. Maksay, Graphical addenda in the technological area of the nodular iron cast rolls production, Acta Polytechnica Hungarica, 5/4 (2008) pp. 15-27

[18] I. Kiss, Multivariate statistical research in areas of the cast hyper-eutectoid steel roll manufacturing, in the melting and alloying processing stages, Acta Polytechnica Hungarica, 17/6 (2020) pp. 41-59

[19] I. Kiss, Research upon the quality assurance of the rolling-mill rolls and the variation boundaries of the chemical composition, Revista de Metalurgia, 44/4 (2008) pp. 335-342

[20] I. Kiss, T. Heput, V. G. Cioata, Some interpretative research in the area of cast iron rolls quality assurance, Acta Metalurgica Slovaca, 16/1 (2010) pp 26-33

[21] I. Kiss, Cast Iron Rolls – An overview on the proper hardness assured by the manufacturing process, Tehnički glasnik–Technical Journal, 13/2 (2019) pp. 92-99

[22] I. Kiss, V. Alexa, The multivariate analysis in area of cast irons with nodular graphite, used in the rolling rolls manufacturing, Machine Design, 4/1 (2012) pp. 43-48

[23] I. Kiss, Focus on quality assurance in the rolls manufacturing-approaches for increasing the rolling-mill rolls qualities, Acta Technica Corviniensis–Bulletin of Engineering, 5/2 (2012) pp. 123-128

[24] I. Kiss, V. Alexa, V. G. Cioată, S. Ratiu, Half-hard cast-iron rolls: statistically research on the manufacturing technology for increase their quality and safety in exploitation, Journal of Physics: IOP Publishing Conference Series, 1781/1 (2021) p. 012054

[25] STAS 9432-85: Rolling mills. Half-hard cast iron rolls specification

Turbine Blade Natural Frequency Estimation Using Various Methods and their Comparisons

Miroslav Spodniak, Jozef Novotňák*, František Heško

Faculty of Aeronautics, Technical University of Košice,
Rampová 7, 041 21 Košice, Slovak Republic

e-mail: miroslav.spodniak@tuke.sk, jozef.novotnak@tuke.sk,
frantisek.hesko@tuke.sk

*Corresponding author

Abstract: The natural frequencies of aircraft jet engine parts are necessary knowledge in the design, safety and operation of the jet engine. The presented article is focused on the area of determining natural frequencies of the turbine blade of the jet engine. The article describes three different methods for determining the natural frequencies of jet engine blades. Acoustic method, method of determining natural frequencies by measuring the vibrations using an accelerometer and determination of natural frequencies by finite element method (FEM) modal analysis. The principles of each method are described in Sections 3 and 4. In Section 5, the achieved results of individual measurements are described. In the conclusion area, Section 6 of this work, the authors describe the results achieved between the various methods and their advantages/disadvantages.

Keywords: natural frequency; turbine blade; frequency measurement; vibration measurement; acoustic measurement; vibrodiagnostic

1 Introduction

Vibration measurement is currently used in many industries and is dealt with by vibration diagnostics, serving as a tool for modern predictive and proactive maintenance methods [1-4]. Vibrations are closely related to the technical condition of the machine, the condition of its parts and their dynamic stress [5-7]. They afford us objective information needed to determine the technical condition of the machines.

In addition to the above options, it is also possible to use vibration diagnostics to determine the natural frequencies of the blades. Natural frequencies are essential part when the stress and life of the jet engine is investigated because the resonance can lead to the failure of the jet engine. The turbine blade is one of the crucial parts of the engine, therefore high reliability has to be ensured during engine operation.

There are several ways to measure the natural frequency of the blade and reveal its frequency spectrum. The most commonly used methods include acoustic measurement or vibration measurement using an accelerometer [8]. In addition to these methods, however, it is possible to determine the natural frequency of the blade also on the basis of calculations.

Nowadays, there is a large number of the various methods for natural frequency estimation [9-15]. The methods can be divided into two separate categories, the first one is mentioned based on experimental methods as the mentioned measurements and second category is numerical method. When using, for example, finite element method (FEM), it is possible to carry out the numerical modal analysis and estimate natural frequencies for the turbine blade for the arbitrary number of modes. The main goal of the article is to present the results of the natural frequencies for iSTC-21v turbine blade numerically estimated and verified by the experimental results and secondly selection of the appropriate method for natural frequency estimation for the mentioned part [16]. An object of the research as mentioned is a turbine blade of iSTC-21v (Figure 1) jet engine, which is an experimental jet engine used mainly for the research purposes.

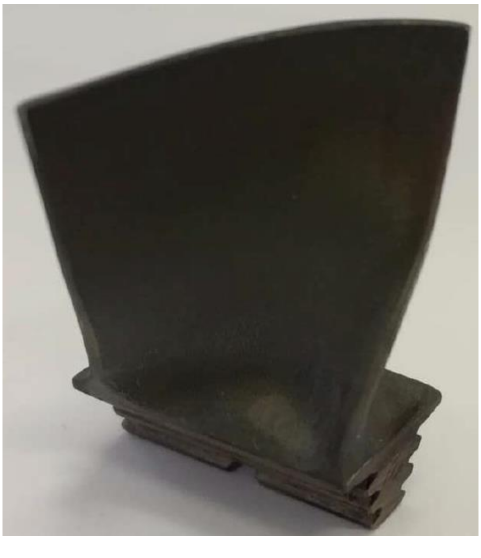


Figure 1
Turbine blade of iSTC-21v jet engine

2 Analysis of Vibrodiagnostic Signal

The vibrodiagnostic signal can be analyzed in the time domain or frequency domain. The vibrations are mostly random in nature and are composed of many frequency components and can be described by amplitude and phase at a given point in time.

2.1 Signal Analysis in the Time Domain and in the Frequency Domain

The analysis of the vibrodiagnostic signal of the time domain (see Figure 2) is based on the evaluation of the parameters of the time courses of the signals. In the time domain, it is easy to evaluate the instantaneous, mean and effective values of the signal or signal envelope. In the case of the predominant random component of the signal (so-called random vibration), selected statistical calculations can be applied for analysis. Time analysis is suitable if there is a single or at least dominant source of vibration, or otherwise there is a loss of diagnostic information in the signal noise caused by the transmission of vibrations from different areas of the machine complex and the possibility of locating the cause of vibration is very limited.

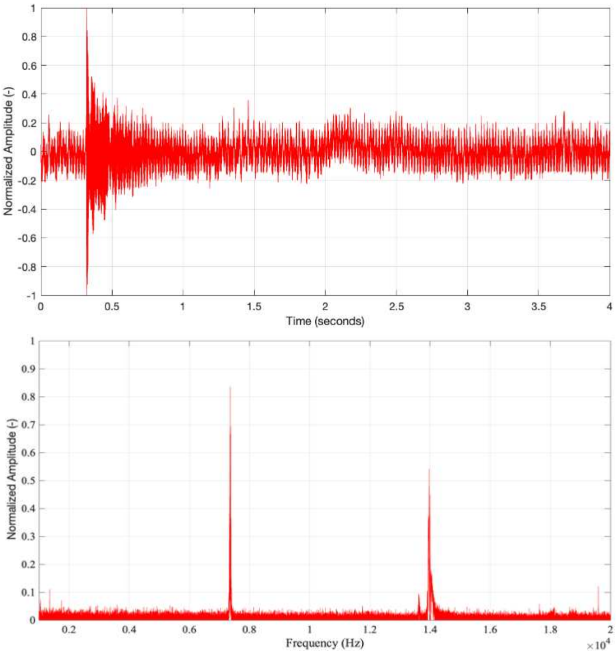


Figure 2

Measured signal in the time domain (top) and in the frequency domain (bottom)

Frequency analysis eliminates the disadvantages of time domain. The aim of spectral analysis is to describe the distribution of signal components in the frequency domain (see Figure 2), to express the analyzed signal using orthogonal basis functions.

A Discrete Fourier Transformation (DFT) can be used to determine the frequency components of the signal. The DFT calculates finite sequences in the time and frequency domains. Several fast algorithms are available for DFT calculation.

Fast Fourier Transformation (FFT) is a typical representative of a fast algorithm for DFT calculation [14]. For a complex input sequence (this is the most general case, the input sequence can also be purely real)

$$x(n) = x_r(n) + jx_i(n) \qquad n = 0,1,\dots,N-1 \tag{1}$$

the FFT algorithm is defined by the equation

$$X(k) = X_r(k) + jX_i(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi kn}{N}} \qquad k = 0,1,\dots,N-1 \tag{2}$$

where, *N* is the number of samples. For inverse FFT (IFFT) we can use equation:

$$x(n) = x_r(n) + jx(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(k)e^{j2\pi\frac{kn}{N}} \qquad n = 0,1,\dots,N-1 \tag{3}$$

All FFT calculations assume a linear system. The length of the classical FFT calculation depends on the length of the input signal, which is the shortest for the lengths of powers of 2. The FFT calculates the individual frequency components of the measured time complex signal, according to predetermined frequency range and resolution requirements. The number of values of the frequency spectrum is halved with respect to the number of values of the time signal, where f_{max} is equal to half of the sampling frequency $f_{vz} = 1/\Delta t$. This is related to Shannon's sampling theorem, according to which the sampling frequency must be at least twice as high (Nyquist frequency) as the frequency of the highest harmonic component contained in the measured signal.

3 Acoustic Frequency Measurement

We used a condenser microphone for acoustic measurement of the blade frequency. The frequency range of the microphone is from 40 Hz to 20 kHz (±5 dB). The signal was sampled with a resolution of 24 bits and a sampling frequency of 48 kHz. The measurement process is shown in the Figure 4. The experimental setup of frequency measurement using a microphone is shown in Figure 8.

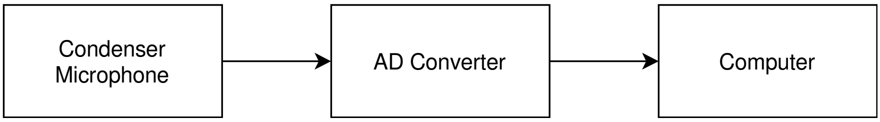


Figure 4
Block diagram of acoustic frequency measurement

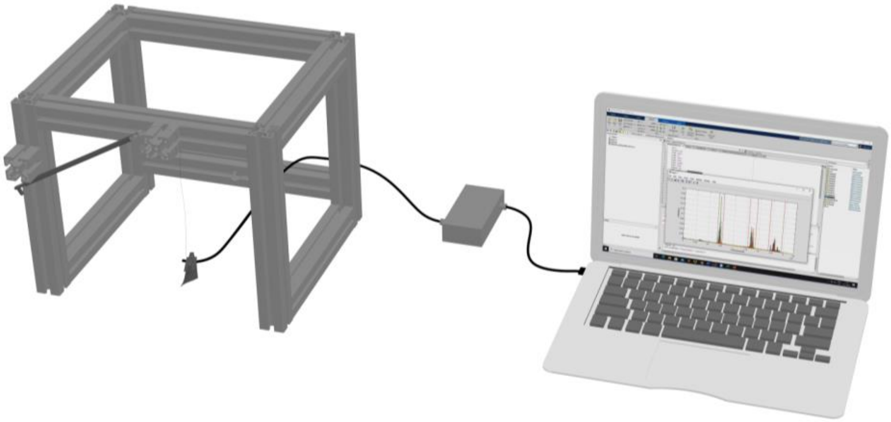


Figure 5
Experimental setup of frequency measurement using a microphone

4 Frequency Measurement Using a Piezoelectric Accelerometer

A piezoelectric accelerometer KD-37 was used to measure the frequency. Its use is suitable, especially in terms of a wide frequency range. The accelerometer generates an electric charge at its output, but its value is very low and it was necessary to amplify it. A charge amplifier was designed (see Figure 6) to amplify the signal, which was adapted exactly for this type of accelerometer, due to its frequency response and electrical parameters (see Table 1).

Table 1
KD-37 Accelerometer parameters

Parameter	**Value**	**Units**
Charge Sensitivity	60	pC/g
Linear frequency range (±3 dB)	15000	Hz
Sensor Capacity	0.61	nF
Cable Capacity 1.5 Meters Long	0.15	nF
Mass of the Accelerometer	45	grams

Given the above parameters the capacitance of the C1 capacitor found in the feedback of the charge amplifier is set to 0.76 nF to obtain a charge amplifier gain of 1 mV/pC. Resistor R1 affect the lower limit frequency and resistor R2 affect the upper limit frequency [18-20]. These frequency limits can be defined with respect to the use of the sensor (for example, the assumed frequency measurement band), but it is also necessary to take into account the sensor parameters (frequency range for example). TL071 was used as an operational amplifier.

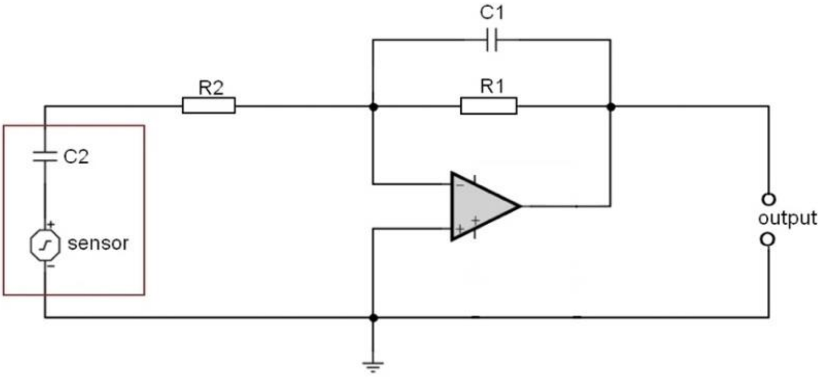


Figure 6
Circuit diagram of a charge amplifier

The measurement process consists of two steps. The first step is to measure the frequency and obtain the necessary data. The second step is the processing and evaluation of the measured data using MATLAB software.

The basic element of frequency measurement is a piezoelectric accelerometer. The output of the accelerometer is fed to the input of the charge amplifier, where the signal is adjusted and amplified. After amplification, the signal is fed to an A/D converter, where it is sampled at a frequency of 48 kHz with a resolution of 24 bits and converted into digital form. After this process, the signal is sent to the computer, where it is recorded and ready for further processing. The whole measurement process is described by a block diagram in Figure 7. Photograph of the experimental setup of frequency measurement using an accelerometer is shown in Figure 8.

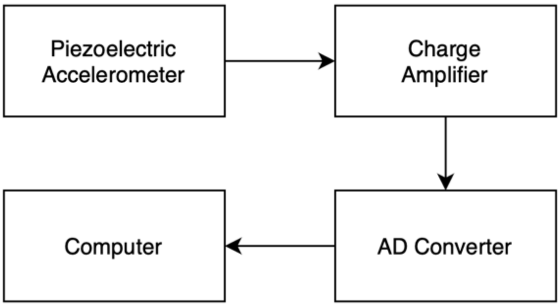


Figure 7
Block diagram of frequency measurement

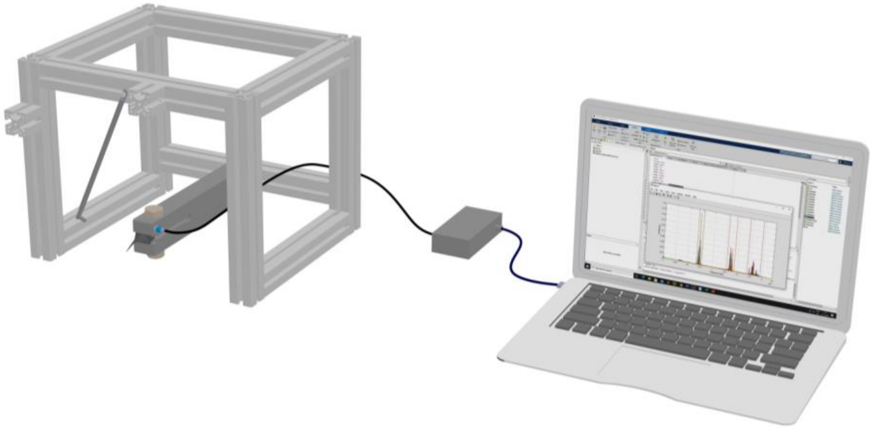


Figure 8
Experimental setup of frequency measurement using an accelerometer

5 Results of Spectral Analysis of the Measured Signal

For data processing of the measured signal, we used MATLAB. Data must be converted from the time domain to the frequency domain. The ***fft*** function [17] was used to detect individual frequencies in the measured signal.

The results obtained by the calculation method were compared with the results of acoustic frequency measurement and with the results of frequency measurement with an accelerometer.

In order to ensure a safe operation, range for jet engine the Campbel diagram is constructed and for the creation of such a diagram, the resonant frequencies are essential part, therefore the research is focused on resonant frequencies and peaks that occurs in the following Figures. Oftentimes, first three modes are important for the Campbel diagram creation. In the proposed article, the first three turbine blade modes for the iSTC-21v jet engine are investigated.

As it was mentioned in previous chapters, natural frequencies are experimentally measured using two methods, or two different sensors, thus acoustic sensor and the accelerometer. The third method for the natural frequency estimation is devoted to the finite element modelling approach and natural frequencies for the turbine blade are computed.

5.1 Acoustic Results

The results of the first mentioned method are presented in Figure 9, where the dependency between the frequency and its amplitude can be seen. A total of 11 turbine blades of the iSTC-21v jet engine were investigated and the measurement for each blade was repeated 5 times. During the experiment each blade was loosely attached using the nylon cord (see Figure 5) and the frequency was excited using mechanical hammer. The first resonant mode of the blade is in a range 6800 – 7500 Hz, a second mode is 13500 – 14500 Hz and the third one is in a frequency range 17000 – 20000 Hz.

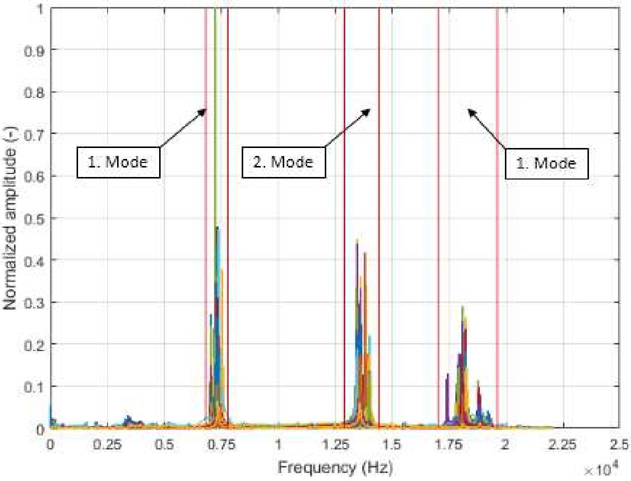


Figure 9
Measured frequency signal for turbine blade using acoustic method

5.2 Accelerometer Method Results

The accelerometer was used for the second experiment, the same number of turbine blade (11) of the iSTC-21v jet engine was measured five times repeatedly for each blade, to ensure the measurement correctness. The excitation was performed similarly like during the acoustic experiment, however, the blade was attached in the clamp to which the accelerometer was attached (see Figure 8). The results of the accelerometer method are presented in Figure 10. When we compare the results with the acoustic measurement, similarity between the two methods is undoubtedly proved. Also, the first resonant mode of the blade is in a same range 6800-7500 Hz, second mode has the range 13500-14500 Hz and the third one is in a frequency range 17500-20000 Hz.

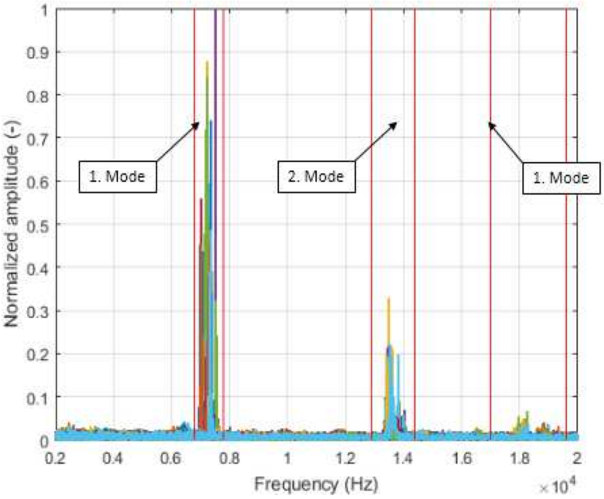


Figure 10
Measured frequency signal for turbine blade using accelerometer

5.3 FEM Modal Analysis Results

The first step when the FEM modal analysis is performed is to create the geometry of an investigated object, in order to perform a modal analysis, the geometry of the turbine blade was created and is described in the article [9]. However, the geometry in the leading edge area was not completely matching the real geometry, therefore for this study the geometry and mesh was updated and the final mesh is shown in Figure 11. The old and updated geometry is shown in Figure 11, the old one is in the left and new updated geometry of the turbine blade is in the right part of the Figure 11. Modal analysis is perfomed in the ANSYS APDL software. The number of finite elements is 368591 and the element type is SOLID 95 hexa elements, the frequency range for the modal analysis is set from 0-24 kHz. The material of the blade is ZS6k alloy. The nodes sets in the fir tree root area are created for the boundary conditions application. Translation in x, y and z direction is fixed. Such a boundary condition is simulating the real mounting conditions in turbine disc during the engine operation. The results of the FEM modal analysis are presented in the following figures, where are computed the shapes and natural frequencies.

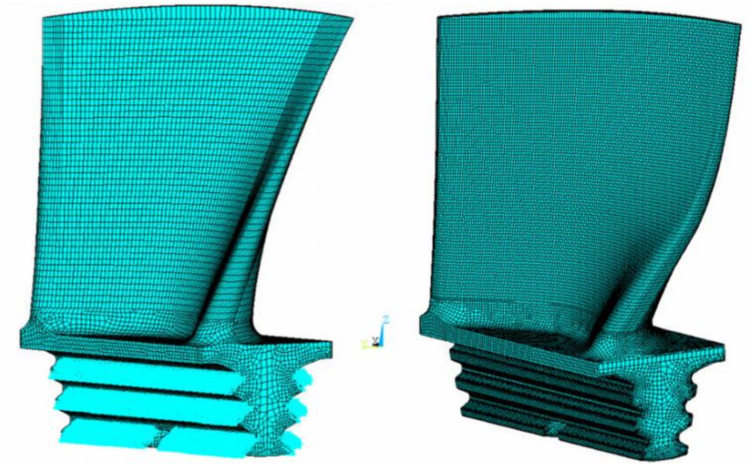
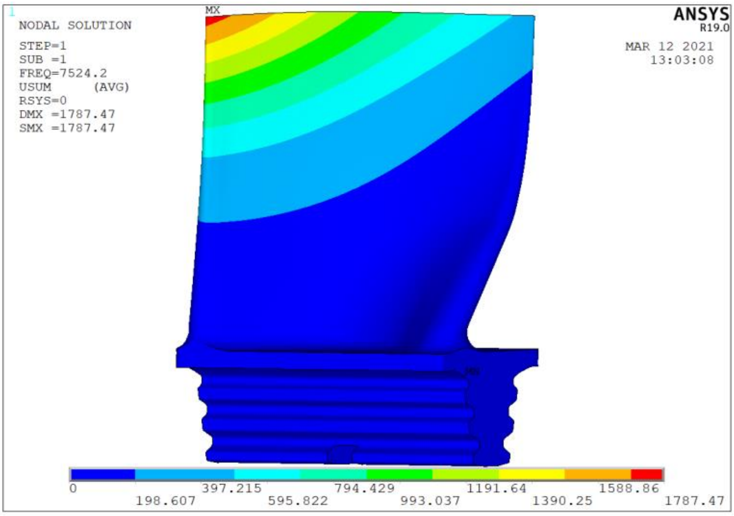


Figure 11

Finite element model for modal analysis in ANSYS APDL

The first mode is presented in Figure 12, natural frequency has a value 7524.2 Hz.



NODAL SOLUTION
STEP=1
SUB =1
FREQ=7524.2
USUM (AVG)
RSYS=0
DMX =1787.47
SMX =1787.47

Figure 12

First natural frequency estimated in ANSYS APDL

Second resonant mode for turbine blade is presented in the Figure 13, where we can see the shape for second natural frequency, which is 13377.8 Hz. The maximal deformation of the turbine blade for the first mode is mainly concentrated in the trailing are at the tip of the blade, however, the deformations for the second mode are concentrated in the leading edge are also at the tip of the blade.

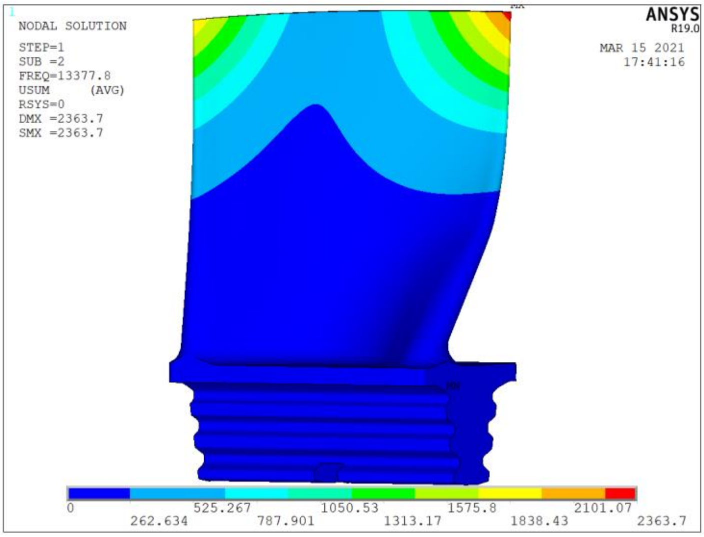


Figure 13
Second natural frequency estimated in ANSYS APDL

As it can be seen in Figures 12 and 13 the natural frequency for both modes is in the same ranges as during the experiment. Third natural frequency estimated using modal analysis in ANSYS APDL has a value 19261.4 Hz and the deformation during this resonant mode are shown in Figure 14. The character of the deformation is in comparison with previous two modes slightly different, the maximal deformation is concentrated in the middle of the blade chord at the tip as can be seen in Figure 14.

An estimated frequencies using FEM method and frequencies measured by mentioned methods are presented in the Table 2.

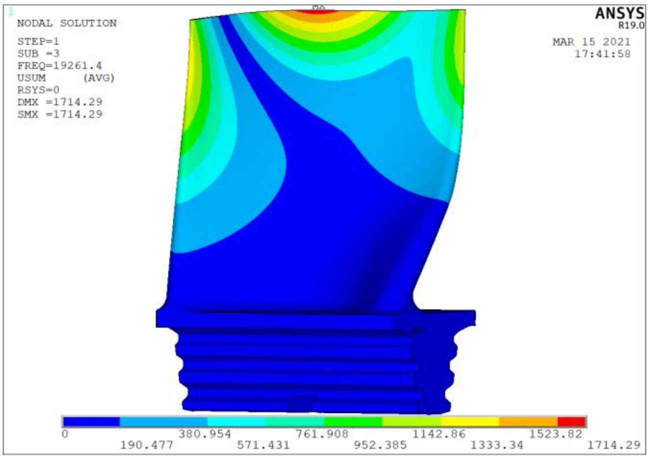


Figure 14

Third natural frequency estimated in ANSYS APDL

Table 2

Turbine blade natural frequencies resume

Method	Measurement	1. Mode [Hz]	2. Mode [Hz]	3 .Mode [Hz]
Acoustic	1	7523	13872	18229
	2	7531	13864	18238
	3	7523	13864	18213
	4	7531	13880	18238
	5	7523	13872	18229
Acoustic Aver.		7526.2	13870.4	18229.4
Accelerometer	1	7528	13860	18250
	2	7527	13850	18245
	3	7528	13855	18255
	4	7528	13880	18260
	5	7527	13860	18248
Accel. Aver.		7527.6	13861.0	18251.6
FEM	1	7524.2	13377.8	19261.4

Apart from the measured frequencies in Table 2, there are also described averages for particular measurements. An average of the acoustic measurement is 7526.2 Hz for the first mode, which is 0.045% higher than the estimated value of the FEM analysis. The second acoustically measured mode is 3.6% higher than calculated frequency and the third one is 5.24% lower value than the frequency calculated by the FEM method. The accelerometer average frequency of the first mode is 7527.6, this value is 0.27% higher than the natural frequency estimated by FEM. The frequency of the second mode is 3.68% higher in comparison with

FEM and the third one is 5.36% lower value than the frequency calculated by the FEM method.

Conclusions

The aim of the article was to compare different methods for determining the natural frequencies of jet engine blades. To determine the frequency, we compared the acoustic method, the measurement method using an accelerometer and FEM modal analysis. When determining the frequency using an accelerometer, we designed our own charge amplifier, adapted to the accelerometer we used.

The results of acoustic and accelerometer measurements were approximately the same in all 3 modes. Compared to the FEM method, we achieved similar results in mode 1 for all 3 measurement methods. In mode 2 and mode 3, the results were slightly different from the FEM method. This can be due to the geometry of the blade and the fact that the blade has already been used in the jet engine. In order to be able to more accurately predict the natural frequencies for modes 2 and 3, it would be necessary to obtain the geometry of the blade already used and include blade wear, which is, however, more time consuming than acoustic or accelerometer measurement.

Comparative measurements have confirmed that the natural frequency measurement can be solved in several ways. With the FEM method, however, it is necessary to create an accurate geometric model of the measured blade and to include blade wear which can be time consuming. Otherwise, if we create only a general blade model that does not involve wearing, then the results at higher frequencies will be inaccurate, which was also confirmed in the above measurements. For acoustic measurement and measurement using an accelerometer, we do not need to solve the geometry of the blade and the determination of the natural frequency can be done faster, but it is necessary to have the necessary equipment. However, the measurement shows that all three methods are applicable and with the correct geometric model, the results are similar.

The issue of determining the natural frequencies of the blades has a wide scope. The natural frequency of the blades will also change, with respect to the mechanical condition of the blade and with respect to the ambient temperature, which could be the subject of further research in the given issue. Further research could determine the trend of the change in the natural frequency of the blade with respect to the ambient temperature, which is also important in terms of safety, in the operation jet engines.

Acknowledgement

This research was funded by the Slovak Research and Development Agency, grant number APVV-17-0184 and grant number APVV-18-0248 and the Research Agency, ITMS code number 313011T557.

References

[1] A. R. Bastami, S. Vahid: “A Comprehensive Evaluation of the Effect of Detect Size in Rolling Element Bearings on the Statistical Features of the Vibration Signal”, in Mechanical Systems and Signal Processing, Vol. 151, 2020

[2] T. A. Shifat, J. W. Hurt: “An Effective Stator Fault Diagnosis Framework of BLDC Motor Based on Vibration and Current Signals”, in IEEE Access, Vol. 8, 2020

[3] M. M. Jovanovič: “Detection of Jet Engine Viper Mk 22-8 Failure in Vibration Spectra”, In: Mitrovic N., Mladenovic G., Mitrovic A. (eds) Experimental and Computational Investigations in Engineering. CNNTech 2020, Lecture Notes in Networks and Systems, Vol. 153, Springer, Cham.

[4] A. S. Komshin and V. I. Pronyakin: “Modern Diagnostic of Aircraft Gas Turbine Engines”, in IOP Conf. Ser.: Mater. Sci. Eng., Vol. 714, 2020

[5] G. Manhertz, A. Bereczky: “STFT Spectrogram Based Hybrid Evaliation Method for Rotating Machine Transient Vibration Analysis”, in Mechanical Systems and Signal Processing, Vol. 154, 2021

[6] S. Fabry, M. Ceskovic: “Aircraft Gas Turbine Engine Vibration Diagnostic”, Magazine of Aviation Development, Vol. 5, 2017, pp. 24-28

[7] Tian Lv and Y. Zhang: "Dynamic Stress Analysis for Vibratory Stress Relief Through the Vibration Platform," 2014 IEEE Workshop on Electronics, Computer and Applications, Ottawa, ON, Canada, 2014, pp. 560-563, doi: 10.1109/IWECA.2014.6845682

[8] Naim Baydar, Andrew Ball: "A Comparative Study of Acoustic and Vibration Signal in Detection of Gear Failures Using Wigner-Ville Distribution," in Mechanical Systems and Signal Processing, Vol. 15, Issue 6, 2001, pp. 1097-1107, doi: 10.1006/mssp.2000.1338

[9] Spodniak, M., Semrád, K., Főző, L., Pavlinský, J.: "FEM Analysis of Natural Frequencies of Jet Engine iSTC-21v Turbine Blade," in SAMI 2019 - IEEE 17th World Symposium on Applied Machine Intelligence and Informatics, Proceedings, January 2019, Article number 8782781, 2019, pp. 287-292

[10] Pridorozhnyi, R. P., Zinkovskii, A. P., Merkulov, V. M. et al: "Calculation-and-Experimental Investigation on Natural Frequencies and Oscillation Modes of Pairwise-Shrouded Cooled Turbine Blades," Strength Mater 51, 2019, pp. 817-827

[11] Gareth L. Forbes, Robert B. Randall: "Estimation of turbine blade natural frequencies from casing pressure and vibration measurements," in Mechanical Systems and Signal Processing, Volume 36, Issue 2, 2013, pp. 549-561

[12] Gillich Gilbert-Rainer, Mituletu Ion Cornel, Nedelcu Dorian, Hamat Codruta Oana: "A procedure for an Accurate Estimation of the Natural Frequencies of Structures," in Vibroengineering PROCEDIA, Vol. 19, 2018, pp. 123-128

[13] M. N. Chekardovskiy, S. M. Chekardovskiy, A. A. Rozboynikov and T. G. Ponomareva: "A Frequency Model of Vibrational Processes in Gas-Turbine Drives of Compressor Stations of Main Gas Pipelines", in IOP Conf. Ser.: Mater. Sci. Eng., Vol. 154, 2016

[14] Yang, Yuan-Jian & Yang, Liang & Wang, Hai-Kun & Zhu, Shun-Peng & Huang, Hong-Zhong,: "Finite Element Analysis for Turbine Blades with Contact Problems," in International Journal of Turbo & Jet-Engines, 2016, pp. 367-371

[15] Poulose, P., Hu, Z.,: "Strength Evaluation and Failure Prediction of a Composite Wind Turbine Blade Using Finite Element Analysis," in Proceedings of the ASME 2010 International Mechanical Engineering Congress and Exposition, Vol. 3, 2010, pp. 295-301

[16] Beneda, K., Andoga, R., Andoga, Főző, L., "Linear Mathematical Model for State-Space Representation of Small Scale Turbojet Engine with Variable Exhaust Nozzle," in Periodica Polytechnica Transportation Engineering, Vol. 46, No. 1, pp. 1-10, 2018

[17] T. Harčarik, J. Bocko and K. Masláková: "Frequency Analysis of Acoustic Signal using the Fast Fourier Transformation in MATLAB", in Procedia Engineering, Vol. 48, 2012, pp. 199-204

[18] J. Novotňák, M. Šmelko and M. Fiľko: "The Design of the Engine Control Unit of Small Aircraft Engine", in Acta Avionica: journal of science, Vol. 20, No. 2, Košice (Slovensko): Faculty of Aeronautics, 2018, pp. 29-37

[19] M. Kellet: "Charge Amplifiers for Piezo Electric Accelerometers" [Online] Available on the Internet: <http://www.mkesc.co.uk/Chargeamps.pdf>

[20] J. Karki: "Signal Conditioning Piezoelectric Sensors " [Online] Available on the Internet: <http://www.ti.com/lit/an/sloa033a/sloa033a.pdf>

Analysis of Network Traversal and Qualification of the Testing Values of Trajectories

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 SZE KVJT and JKK; 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: This research work is aimed directly at the study of network traversal, for the design of vehicle dynamics and driving test programs. The work can be applied more widely to the qualification of test tracks. In addition to general modelling, this method can also be used to investigate the formation of loops in order to take into account, the sub-routes, multiple times. An important area of its application is the more comprehensive analysis of complex loads, as well as, the development of learning algorithms, which can be achieved by repeating multiple traversals of certain track sections within a series of measurements, and can be used for the development of on-board vehicle systems. The mathematical modelling is presented through the application of the geometric graph and subgraphs of the track. The properties of the Markov model extracted from the connection matrix of the large-scale network model are also presented. In this way, the modelling is extended to the application of the connection matrix of the large-scale network model as well. The modelling and computational details are demonstrated by means of a computer based, algebraic program. This modelling and the results of the calculations, will allow the further development of a test program design and related evaluation methods.

Keywords: Study of network traversal; mathematical modelling; connection matrix; Markov model; computer algebra; test program design; evaluation method

1 Introduction

This research work is directly related to the study of network traversal and can be applied to the design of vehicle dynamics and driving test programs, and more broadly to the qualification of test tracks.

Within the research tasks related to the vehicle test track, the analysis of track capabilities is a key area. The work serves as an analysis of track capabilities for testing autonomous vehicles and results in a new and efficient method.

When testing autonomous vehicles [15, 18, 22, 24], both dynamic track elements and urban track elements are important to ensure that together they represent the complex effects of the real environment at the highest level.

Reducing the complexity of the calculations for this complex problem is also of great importance for our investigations.

Therefore, we propose a method that either (I) compiles an optimal test program for the existing test track under a set of criteria and selects the set of best trajectories for the given objective, or (II) determines the set of best trajectories for the given objectives in the design phase of the tracks.

The latter can provide direction for redesign and the procedure also provides algorithmic methods for designing the structure of the track models.

2 Studies on the Design of Test Tracks

The development of autonomous driving technologies is receiving considerable attention, extensive research and validation is carried out to test the autonomy and safety of vehicle systems, I. Passchier, G. v. Vugt, and M. Tideman [7]. The test track is designed and constructed to simulate the real environment along with the foreseeable risks, R. Chen, M. Arief, D. Zhao [2], so in the studies, the trajectories must adapt faithfully and flexibly to reality. These benefits have already led to the creation of several world-class CAVs, e.g. T. Stevens. [14], Harman [3], Mcity Test Facility, [4], D. Muoio. [5], ACM. Project [1]. However, the proof of the CAV evaluation theory in design is not yet widely reported in the literature. This was pointed out and a mathematical model was developed by R. Chen, M. Arief, D. Zhao [2]. Indeed, the CAV's ability to operate in the spatial, field conditions that are regularly encountered by autonomous vehicles in real-world situations is important. These include compliance with road rules, in addition to reacting to other vehicles and road users, as well as to frequently encountered hazards.

Examination of such cases leads to the capability checklist that can be used as an evaluation criterion for CAV systems, C. Nowakowski, S. E. Shladover, C.-Y. Chan, and H.-S. Tan. [6], Torc. [16].

First, R. Chen, M. Arief, D. Zhao [2] assessed a systematic approach that uses an optimal model with the aim of maximizing the testing ability of the proof of CAV evaluation.

The goal of the design approach is to map scenarios in a given space that maximize the evaluation capability of the CAV. The research is important from two perspectives: 1) It is not clear how typical scenarios are currently classified and selected from a large-scale real-world driving dataset; 2) It is currently unclear how these scenarios and proofs can be mapped.

In order to enable a particular scenario suitable for testing in their task set, they had to draw maps of a set of roads and intersections that support the accurate implementation of the concurrence of these road elements in evaluation planning. In addition, a so-called “value measure” can also be defined for it, which can be presented and used for the evaluation of road use. It is a very important finding that **the value of road assets supports versatility, multi-scenario usability, and also supports the universality of scenarios.**

Modelling considerations, constraints:

> When measuring the value of road assets **S**, the pre-processed sequence **A** = {a_1, a_2, …, a_{NA}} can be used as a basis in the limited construction area.
>
> 1^0. Road elements in one asset shall not overlap roads in other assets.
>
> 2^0. The extent of road assets shall be contained within the constrained space **S**.

In Phase 1, a set representing the highest complexity asset value is selected from a subset of the feasible road sections (from all pre-generated road data sets).

In Phase 2, the pre-selected set is optimized according to **S** in the given space, taking into account the elasticities for the transitions in the scenarios.

The aim of the two-phase approach is the following:

> ***Phase 1 is thus a model for selecting the most valuable feasible subset***, taking into account constraints in a strict sense.
>
> ***Phase 2: the optimization model that*** **takes into account the pre-selected optimal set where the objective,** ***based on Phase 1,*** **is to be maximized both in number and in the value of the elements.**

In the above, the constraint ensures that for each connected asset pair, at least one transition road connecting the boundary nodes must be selected. The constraint ensures that any designated temporary road candidate does not access another internal segment unless the case occurs that a node is constructed. Regarding the structure of the model, based on the above, it is the basis of a design framework that provides a geographic map for the implementation of a wide variety of CAV scenarios in the constrained space and the scenario provides maximum flexibility for transitions.

However, the proposed valuable method ran into a serious computational problem. Based on the prescribed model, the defined optimization procedure leads to an NP problem, mainly due to the binary tools used, which ensure the validity of the asset forms and combined patterns, and some constraint conditions are not convex either. The use of an excessive number of binary variables and non-linear constraints in this method required a significant amount of branching processes and relaxation solutions, which became the largest computational burden in the studies. Due to the resulting computational complexity, further work is needed to address the problems, according to the authors, focusing on computational solutions by breaking down the problem.

3 The New Modelling Considerations and Methods

As seen in the previous chapter, the “Xcity” study discusses a new design method for the construction of test tracks based on a complex optimization calculation. In this regard, the authors presented a design procedure using a limited number of road assets. However, the method is not applicable in practice due to its exponential computational complexity.

In our case, the “ZalaZone” test track [17], on the other hand, is already a planned and implemented test track, which is already partially operational. This test track will provide a complete test environment and multi-level testing capability for future vehicles and communication technologies, from prototype testing to series product development.

Within the framework of the first phase, the dynamic surface, the high-speed handling track, the first part of the Smart City Zone was implemented; this is followed by the completion of the motorway section, various highway sections, and the Smart City. By early 2022, the low-speed handling track, the ramps, the noise measurement surface, and the ADAS test surface will be completed.

In this case, the real track curves are composed of real sector pieces, thus real curve pieces and sectors are the building blocks of trajectories, i.e. test routes. It is very important that in this way the network approach can be perfectly applied in modelling as well [19-21] [23]. Then, taking into account all potential branch points, all possible real routes of the test track can be generated by considering each distribution point. With the algorithm we present, real test routes can be generated at high speed without any geometric and hardware-related computational constraints. The generation of all possible track trajectories, for this purpose we use the large-scale network model we have developed, in particular its connection matrix [8-13]. To determine all possible trajectories of the track, we introduce the distribution marker $\alpha_{i,j}$, which accurately indicates all routes that originate from the branch it induces. In the sample track shown in Figure 1, we consider cases A and B, which use different sector controls in each sector. Thus, different closed-curve trajectories can be constructed for testing purposes in the case of A and B and the evaluation methods can also be examined. The aim is to identify the most valuable closed track trajectories that best meet the tester's criteria.

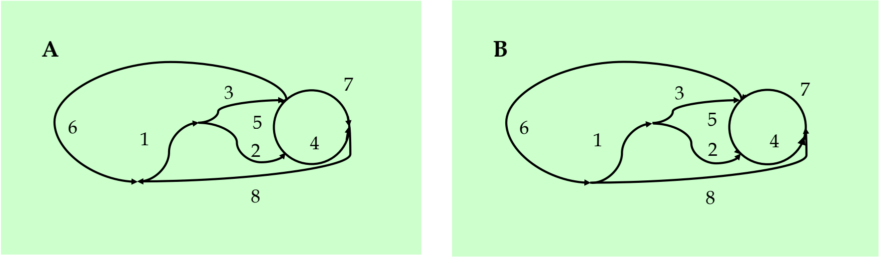


Figure 1

In the case of the track consisting of the sub-track curves Nos. 1-8, two different sets of trajectories, one of type A and one of type B, are created for the traversal due to the different sector controls

Thus, our research aims to develop applicable mathematical methods that accurately take into account the geometry of the trajectories and are also able to perform an exact qualification-evaluation procedure for the trajectories. Our goal is that the matrix transformation expansion method to be developed for generating trajectories should provide fast, real-time computational performance, and thus be efficient in this respect as well, as opposed to the Xcity method, which requires NP computational power.

4 Selection of the Optimal Trajectories

The possibility arises for the optimal selection of the trajectories consisting of the connected curve sections with sequentially structured dynamic programming. In this case, to perform the sequential optimization, e.g. Bellman's principle of optimality can be applied. The approach is to apply the principle of optimality to discrete, deterministic systems, taking into account that "an optimal policy can only consist of optimal sub-policies".

The reason for not doing so in our case is justified by the fact that we can define a very elegant and fast matrix transformation expansion method based on the traffic model under discussion. In terms of traffic on the test track, the connection matrix $\boldsymbol{K}_{11}$ contains the v_{ij} dynamic speed connections between the internal sector elements. For the test track consisting of *n* internal sectors (*i,j* = 1,2,...*n*), in the case of an arbitrary sector *"j"* the connection matrix ***K***$_{11}$ determines to which sector or sectors *"i"* the controlled amount of material flows at an also controlled speed. To solve our problem, we form the transition probability or distribution matrix **P** from the matrix **K**$_{11}$ by keeping only the $\alpha_{i,,j}$ distribution values for the matrix elements and neutralizing the effect of the other dynamic control functions by replacing them with a value of 1. In addition, it is also necessary to replace the elements in the main diagonal with values of 0, so that the result of all considered further steps does not disappear (i.e. it remains fixed). The effects of material removals occurring in real dynamic processes are not relevant in this study, because the task is only to determine all the possibilities of further steps, at the same time it is necessary to mark and preserve all previous locations in order to reconstruct all routes. Our basic concept is that all previous splits will leave an accurate trace along the routes that resulted from the split. So **markers** determine the formation of each route. These well-identifiable markers are the $\alpha_{i,,j}$ distributions. For the above, we use the Markov property derived from the matrix **K**$_{11}$. For each *j*-th column of the matrix (j=1,2, ... , *n*), the $\alpha_{ij} \geq 0$ discrete distribution (the sum of these in column *j* is 1) depending on the condition of staying in sector *j,* determines the probability of moving from sector *j* to sector *i*: (In our case: *i≠j*)

$$\alpha_{ij} = P(i \mid j) \qquad (1)$$

Thus, the sum of the column elements of the matrix **P=P[** $\alpha_{i,,j}$**]** is: $\sum_{i=1}^{n} \alpha_{ij} = 1$ (j=1,2, ... , n). It can be clearly seen that the matrix **P** written in this way defines a discrete Markov chain. Note that since the $\alpha_{ij}(x(t),t)$ values can be considered constant only for short periods of time in a real traffic network and actually depend on time and also on the vehicle density state characteristic *x(t)* of the sectors, the *Markov chain defined here is inhomogeneous*. The elements of the matrix **P** cannot be said to be positive definite either, since there are a large number of 0 elements among them, so the *Markov chain under discussion is also irregular*. It is important to emphasize that when defining paths, distributions are not time- and state-dependent features. These are fixed constants because we want to consider all possible road sections in the applied procedure. The analysis is then carried out by starting from the input sector No. 1, considered as the gate of the track, and determining the distributions (scatterings) proceeding step by step. Accordingly, at the initial time, we are only in sector No. 1 with a probability of p>0. The probabilities of the initial states "1" are determined by the vector **p**$_1$: according to the conditions, the probability of staying in sector No. 1 is p, while the probabilities of staying in the other sectors take the value of 0:

$$p_1 = \begin{bmatrix} p \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{2}$$

The following state probabilities (3) are determined by the vector $\mathbf{p}_2$ calculated using the matrix **P** and the vector $\mathbf{p}_1$, where the matrix **P** contains the distributional directions and values for proceeding from sector No. 1:

$$p_2 = P \cdot p_1 \tag{3}$$

The following vector $\mathbf{p}_3$, which determines the state probabilities (4), is calculated in a similar recursive way using the matrix **P** and the vector $\mathbf{p}_2$. In this case, the matrix **P** also contains distributional directions and values for proceeding from sector No. 2:

$$p_3 = P \cdot p_2 \tag{4}$$

Finally, the vector $\mathbf{p}_\mathrm{n}$ determining the *n*-th state probabilities is also calculated recursively using the matrix **P** and the vector $\mathbf{p}_{\mathrm{n-1}}$:

$$p_n = P \cdot p_{n-1} \tag{5}$$

Referring to the relation for discrete Markov chains, it can be clearly seen from the above derivation that the state probability vector $\mathbf{p}_\mathrm{n}$ can be generated by a one-step method as the product of the *n*-th power of the matrix **P** and the vector $\mathbf{p}_1$:

$$p_n = P^n \cdot p_1 \tag{6}$$

After *n* steps, if the sectors in the domains have no outflow end, i.e. there is no further transfer, and loops can be applied to the routes but not in infinite cycles, then after a finite number of steps a steady state occurs, thus the probabilities of staying in each sector are no longer modified by further application of the algorithm:

$$\mathbf{p}_\mathrm{n} = \mathbf{p}_{\mathrm{n+1}} = \mathbf{p}_{\mathrm{n+2}} = \ldots = \mathbf{p}_{\mathrm{n+k}} = \ldots \tag{7}$$

The significance of this is that the non-zero coordinates of the vector $\mathbf{p}_\mathrm{n}$ can be used to determine how many different routes, which can be considered parallel, have led from the input sector No. 1 to any other sector. In this context, the routes that are most important are those that lead to the sectors consisting of the most elements, which can be considered as the most distant "outputs", as these can provide the most opportunities for a qualitative analysis. This will be illustrated by the examples shown in Figures 1a and 1b.

5 Presentation of the Method for Two Types of Test Tracks Consisting of 8 Sectors

The method plays an important role in the machine-based generation of all trajectories in the case of a test track after taking an arbitrary input. It can play an equally important role in the definition and evaluation of various criteria and rankings according to the test criteria. Let us consider the "test track" consisting of n=8 sectors as seen in Figures 1a and 1b, and all possible closed-curve trajectories that can be constructed from a selected sector on track A. In the case of sector No. 1 that means those, which start from and arrive at sector No. 1, while in the case of track B those starting from sector No. 1 and ending in sectors Nos. 6, 7, and 8.

Let us consider the transition probability matrices **P** of the test tracks shown in Figures 1a and 1b (8a) and (8b), which are the matrices defining the relationship between the different elements *i* and *j*.

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{2,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{3,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \alpha_{4,2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \alpha_{5,2} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_{6,3} & 0 & \alpha_{6,5} & 0 & 0 & 0 \\ 0 & 0 & \alpha_{7,3} & 0 & \alpha_{7,5} & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_{8,4} & 0 & 0 & \alpha_{8,7} & 0 \end{bmatrix} \qquad (8a)$$

$$\mathbf{P} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{2,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \alpha_{3,1} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \alpha_{4,2} & 0 & 0 & \alpha_{4,5} & 0 & 0 & 0 \\ 0 & 0 & \alpha_{5,3} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \alpha_{6,3} & 0 & 0 & 0 & \alpha_{6,7} & 0 \\ 0 & 0 & 0 & \alpha_{7,4} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \alpha_{8,6} & 0 & 0 \end{bmatrix} \qquad (8b)$$

According to the construction, the probability of staying in sector No. 1, which is considered as input, is initially p, and the probabilities of staying in the other sectors take the value of 0. This initial state is defined by the vector P1:

$$\mathrm{P1} = \begin{bmatrix} p \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Calculating the probabilities of staying in the sectors as we proceed, respectively: $p_2 = P \cdot p_1$, $p_3 = P \cdot p_2$, ..., $p_n = P \cdot p_{n-1}$, we obtain the following sequence of vectors, which determine the state probabilities for tracks A and B:

$$\mathrm{P2} = \begin{bmatrix} p \\ \alpha_{2,\,1}\, p \\ \alpha_{3,\,1}\, p \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad (9a) \qquad\qquad \mathrm{P2} = \begin{bmatrix} p \\ \alpha_{2,\,1}\, p \\ \alpha_{3,\,1}\, p \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad (9b)$$

Where, index “a” refers to track A, while index “b” refers to track B.

During the calculation, after the 5th transformation step (which has already determined all possible routes), the probability vector P_j does not change any more. This is due to the fact that the transition probability matrix used was written in such a way that in case “A”, sectors Nos. 6 and 8 do not pass vehicles to any other sector any more, while in case “B”, the routes end in sectors Nos. 6, 7 and 8.

All routes starting from input sector No. 1 can be specified by determining how many routes lead to each of the end sectors separately. (In case "A" from No. 1 to outputs Nos. 6 and 8; in case "B" from No. 1 to outputs Nos. 6, 7 and 8.)

For case “A”: For the routes leading from sector No. 1 to output sector No. 6, the 6th coordinate of the vector must be examined. In this regard, we see two sums resulting from the fact that we reached No. 6 on two parallel routes.

$$\mathrm{P3} = \begin{bmatrix} p \\ \alpha_{2,\,1}\,p \\ \alpha_{3,\,1}\,p \\ \alpha_{4,\,2}\,\alpha_{2,\,1}\,p \\ \alpha_{5,\,2}\,\alpha_{2,\,1}\,p \\ \alpha_{6,\,3}\,\alpha_{3,\,1}\,p \\ \alpha_{7,\,3}\,\alpha_{3,\,1}\,p \\ 0 \end{bmatrix} \qquad (10a)$$

$$\mathrm{P3} = \begin{bmatrix} p \\ \alpha_{2,\,1}\,p \\ \alpha_{3,\,1}\,p \\ \alpha_{4,\,2}\,\alpha_{2,\,1}\,p \\ \alpha_{5,\,3}\,\alpha_{3,\,1}\,p \\ \alpha_{6,\,3}\,\alpha_{3,\,1}\,p \\ 0 \\ 0 \end{bmatrix} \qquad (10b)$$

$$\mathrm{P4} = \begin{bmatrix} p \\ \alpha_{2,\,1}\,p \\ \alpha_{3,\,1}\,p \\ \alpha_{4,\,2}\,\alpha_{2,\,1}\,p \\ \alpha_{5,\,2}\,\alpha_{2,\,1}\,p \\ \alpha_{6,\,3}\,\alpha_{3,\,1}\,p + \alpha_{6,\,5}\,\alpha_{5,\,2}\,\alpha_{2,\,1}\,p \\ \alpha_{7,\,3}\,\alpha_{3,\,1}\,p + \alpha_{7,\,5}\,\alpha_{5,\,2}\,\alpha_{2,\,1}\,p \\ \alpha_{8,\,4}\,\alpha_{4,\,2}\,\alpha_{2,\,1}\,p + \alpha_{8,\,7}\,\alpha_{7,\,3}\,\alpha_{3,\,1}\,p \end{bmatrix} \qquad (11a)$$

$$\mathrm{P4} = \begin{bmatrix} p \\ \alpha_{2,1}\, p \\ \alpha_{3,1}\, p \\ \alpha_{4,2}\,\alpha_{2,1}\, p + \alpha_{4,5}\,\alpha_{5,3}\,\alpha_{3,1}\, p \\ \alpha_{5,3}\,\alpha_{3,1}\, p \\ \alpha_{6,3}\,\alpha_{3,1}\, p \\ \alpha_{7,4}\,\alpha_{4,2}\,\alpha_{2,1}\, p \\ \alpha_{8,6}\,\alpha_{6,3}\,\alpha_{3,1}\, p \end{bmatrix} \tag{11b}$$

$$\mathrm{P5} = \begin{bmatrix} p \\ \alpha_{2,1}\, p \\ \alpha_{3,1}\, p \\ \alpha_{4,2}\,\alpha_{2,1}\, p \\ \alpha_{5,2}\,\alpha_{2,1}\, p \\ \alpha_{6,3}\,\alpha_{3,1}\, p + \alpha_{6,5}\,\alpha_{5,2}\,\alpha_{2,1}\, p \\ \alpha_{7,3}\,\alpha_{3,1}\, p + \alpha_{7,5}\,\alpha_{5,2}\,\alpha_{2,1}\, p \\ \alpha_{8,4}\,\alpha_{4,2}\,\alpha_{2,1}\, p + \alpha_{8,7}\left(\alpha_{7,3}\,\alpha_{3,1}\, p + \alpha_{7,5}\,\alpha_{5,2}\,\alpha_{2,1}\, p\right) \end{bmatrix} \tag{12a}$$

$$\mathrm{P5}= \begin{vmatrix} p \\ \alpha_{2,1}\,p \\ \alpha_{3,1}\,p \\ \alpha_{4,2}\,\alpha_{2,1}\,p+\alpha_{4,5}\,\alpha_{5,3}\,\alpha_{3,1}\,p \\ \alpha_{5,3}\,\alpha_{3,1}\,p \\ \alpha_{6,3}\,\alpha_{3,1}\,p \\ \alpha_{7,4}\left(\alpha_{4,2}\,\alpha_{2,1}\,p+\alpha_{4,5}\,\alpha_{5,3}\,\alpha_{3,1}\,p\right)+\alpha_{7,8}\,\alpha_{8,6}\,\alpha_{6,3}\,\alpha_{3,1}\,p \\ \alpha_{8,6}\,\alpha_{6,3}\,\alpha_{3,1}\,p \end{vmatrix} \qquad (12b)$$

$$\mathrm{P6}= \begin{bmatrix} p \\ \alpha_{2,1}\,p \\ \alpha_{3,1}\,p \\ \alpha_{4,2}\,\alpha_{2,1}\,p \\ \alpha_{5,2}\,\alpha_{2,1}\,p \\ \alpha_{6,3}\,\alpha_{3,1}\,p+\alpha_{6,5}\,\alpha_{5,2}\,\alpha_{2,1}\,p \\ \alpha_{7,3}\,\alpha_{3,1}\,p+\alpha_{7,5}\,\alpha_{5,2}\,\alpha_{2,1}\,p \\ \alpha_{8,4}\,\alpha_{4,2}\,\alpha_{2,1}\,p+\alpha_{8,7}\left(\alpha_{7,3}\,\alpha_{3,1}\,p+\alpha_{7,5}\,\alpha_{5,2}\,\alpha_{2,1}\,p\right) \end{bmatrix} \qquad (13a)$$

$$\mathrm{P6}=\left|\begin{array}{c} p \\ \alpha_{2,1}p \\ \alpha_{3,1}p \\ \alpha_{4,2}\alpha_{2,1}p+\alpha_{4,5}\alpha_{5,3}\alpha_{3,1}p \\ \alpha_{5,3}\alpha_{3,1}p \\ \alpha_{6,3}\alpha_{3,1}p \\ \alpha_{7,4}\left(\alpha_{4,2}\alpha_{2,1}p+\alpha_{4,5}\alpha_{5,3}\alpha_{3,1}p\right)+\alpha_{7,8}\alpha_{8,6}\alpha_{6,3}\alpha_{3,1}p \\ \alpha_{8,6}\alpha_{6,3}\alpha_{3,1}p \end{array}\right| \qquad (13b)$$

Test routes are shown by the right-to-left readings of the second indexes of the $\alpha_{i,j}$ distributions.

$$\mathbf{P8}[6]=\alpha_{6,3}\alpha_{3,1}p+\alpha_{6,5}\alpha_{5,2}\alpha_{2,1}p \qquad (14)$$

The first route, denoted by I, leads from No. 1 to No. 3 and from No. 3 to No. 6;

The second route, denoted by II, leads from No. 1 to No. 2, then from No. 2 to No. 5 and from No. 5 to No. 6.

The routes leading from No. 1 to No. 8 are similarly shown by the 8th coordinate of the vector. Here we see the sum of 3 members as follows, so we can get from No. 3 to No. 8 on 3 different, parallel routes.

$$\mathbf{P8}[8]=\alpha_{8,4}\alpha_{4,2}\alpha_{2,1}p+\alpha_{8,7}\left(\alpha_{7,3}\alpha_{3,1}p+\alpha_{7,5}\alpha_{5,2}\alpha_{2,1}p\right) \qquad (15)$$

The first route, denoted by III, leads from No. 1 to No. 2, then from No. 2 to No. 4, and finally, from No. 4 to No. 8.

The second route, denoted by IV, leads from No. 1 to No. 3, then from No. 3 to No. 7, and finally, from No. 7 to No. 8.

The third route, denoted by V, leads from No. 1 to No. 2, then from No. 2 to No. 5, from No. 5 to No. 7, and finally, from No. 7 to No. 8.

It can be clearly seen that formulas (14) and (15) obtained by the computer-algebraic method are parametric mathematical formulas. Their evaluation for determining the route codes is done by processing strings and characters. These formulas consist of sums of products and each product is separated from the next by a "+" sign. These products form the strings in which, when analyzing the characters from right to left, only the indices of the alpha character are collected and two identical indices following each other are considered only once.

The sequence of numbers thus determined consists of the serial numbers of the consecutive and interconnected sector elements that form the route. (See routes I, II, ... , V.)

In case "B", similarly to the above, all routes starting from sector No. 1 and ending in either sector No. 6, No. 7 or No. 8 are defined:

$$\mathbf{P8}[6] = \alpha_{6,3}\,\alpha_{3,1}\,p \tag{16}$$

$$\mathbf{P8}[7] = \alpha_{7,4}\left(\alpha_{4,2}\,\alpha_{2,1}\,p + \alpha_{4,5}\,\alpha_{5,3}\,\alpha_{3,1}\,p\right) + \alpha_{7,8}\,\alpha_{8,6}\,\alpha_{6,3}\,\alpha_{3,1}\,p \tag{17}$$

$$\mathbf{P8}[8] = \alpha_{8,6}\,\alpha_{6,3}\,\alpha_{3,1}\,p \tag{18}$$

I: leads from No. 1 to No. 3 and from No. 3 to No. 6; (route defined by P8 [6])

II: leads from No. 1 to No. 3, then from No. 3 to No. 6, from No. 6 to No. 8, and finally, from No. 8 to No. 7; (route defined by P8 [7])

III: leads from No. 1 to No. 3, then from No. 3 to No. 5, from No. 5 to No. 4, and finally, from No. 4 to No. 7; (route defined by P8 [7])

IV: leads from No. 1 to No. 2, then from No. 2 to No. 4, and finally, from No. 4 to No. 7; (route defined by P8 [7])

V: leads from No. 1 to No. 3, then from No. 3 to No. 6, and finally, from No. 6 to No. 8; (route defined by P8 [8])

6 Presentation of Matrix Operations for the Calculations

In summary, in both cases a total of 5 different routes led from input sector No. 1 so that each of the sectors used in a route was only considered once. Based on the above matrix transformation algorithm, all trajectories were determined for the test tracks shown in *Figs. 1a and 1b.* These trajectories, *i*=I., II., … , V, are encoded in the rows of the matrix **Tr** (19a and 19b) in such a way that where 1 is in the *i*-th row, that sector *j* is part of the *i*-th trajectory:

$$\mathbf{Tr} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix} \tag{19a}$$

$$\mathbf{Tr} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 \end{bmatrix} \tag{19b}$$

The elements [i,j] of the matrix **Ck** (20a and 20b) contain the values of the individual trajectory elements (there are 8 sector elements) for the given test program.

$$\mathbf{Ck} = \begin{bmatrix} 1.5 & 2.5 & 3.2 & 5.7 & 4.1 & 7.6 & 3.8 & 9.5 \\ 1.5 & 2.5 & 3.2 & 5.7 & 4.1 & 7.6 & 3.8 & 9.5 \\ 1.5 & 2.5 & 3.2 & 5.7 & 4.1 & 7.6 & 3.8 & 9.5 \\ 1.5 & 2.5 & 3.2 & 5.7 & 4.1 & 7.6 & 3.8 & 9.5 \\ 1.5 & 2.5 & 3.2 & 5.7 & 4.1 & 7.6 & 3.8 & 9.5 \end{bmatrix} \tag{20a}$$

$$\mathbf{Ck} = \begin{bmatrix} 1.5 & 2.5 & 3.2 & 5.7 & 4.1 & 7.6 & 3.8 & 9.5 \\ 1.5 & 2.5 & 3.2 & 5.7 & 4.1 & 7.6 & 3.8 & 9.5 \\ 1.5 & 2.5 & 3.2 & 5.7 & 4.1 & 7.6 & 3.8 & 9.5 \\ 1.5 & 2.5 & 3.2 & 5.7 & 4.1 & 7.6 & 3.8 & 9.5 \\ 1.5 & 2.5 & 3.2 & 5.7 & 4.1 & 7.6 & 3.8 & 9.5 \end{bmatrix} \tag{20b}$$

The elements [i,j] of the matrix **Cv** (21a and 21b) give the values of the sector change for each trajectory element for the given test program. The value is assigned to the transferring sector.

$$\mathbf{Cv} = \begin{bmatrix} 0 & 0 & 0.5 & 0 & 0 & 0.1 & 0 & 0 \\ 0 & 0.5 & 0 & 0 & 1.6 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 1.2 & 0 & 0 & 0 & 1.3 \\ 0 & 0 & 0.5 & 0 & 0 & 0 & 0.2 & 1.1 \\ 0 & 0.5 & 0 & 0 & 1.6 & 0 & 1.2 & 1.3 \end{bmatrix} \tag{21a}$$

$$\mathbf{Cv} = \begin{bmatrix} 0 & 0 & 0.7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0 & 0 & 0 & 0 & 0.2 \\ 0 & 0 & 0.9 & 0 & 0.1 & 0.1 & 0 & 0 \\ 0 & 0.5 & 0 & 0.1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.7 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \tag{21b}$$

The elements [i,j] of the matrix **C** (22a and 22b) are the summed values for each trajectory element for the given test program:

$$\mathbf{C} = \mathbf{Ck} + \mathbf{Cv} = \begin{bmatrix} 1.5 & 2.5 & 3.7 & 5.7 & 4.1 & 7.7 & 3.8 & 9.5 \\ 1.5 & 3.0 & 3.2 & 5.7 & 5.7 & 8.1 & 3.8 & 9.5 \\ 1.5 & 3.0 & 3.2 & 6.9 & 4.1 & 7.6 & 3.8 & 10.8 \\ 1.5 & 2.5 & 3.7 & 5.7 & 4.1 & 7.6 & 4.0 & 10.6 \\ 1.5 & 3.0 & 3.2 & 5.7 & 5.7 & 7.6 & 5.0 & 10.8 \end{bmatrix} \qquad (22a)$$

$$\mathbf{C} = \mathbf{Ck} + \mathbf{Cv} = \begin{bmatrix} 1.5 & 2.5 & 3.7 & 5.7 & 4.1 & 7.7 & 3.8 & 9.5 \\ 1.5 & 3.0 & 3.2 & 5.7 & 5.7 & 8.1 & 3.8 & 9.5 \\ 1.5 & 3.0 & 3.2 & 6.9 & 4.1 & 7.6 & 3.8 & 10.8 \\ 1.5 & 2.5 & 3.7 & 5.7 & 4.1 & 7.6 & 4.0 & 10.6 \\ 1.5 & 3.0 & 3.2 & 5.7 & 5.7 & 7.6 & 5.0 & 10.8 \end{bmatrix} \qquad (22b)$$

The elements [i, j] of the matrix **C•Tr** (23a and 23b) are the test values of the sectors that form the actual trajectories, which product is the Hadamard product of the two matrices:

$$\mathbf{C} \cdot \mathbf{Tr} = \begin{bmatrix} 1.5 & 0. & 3.7 & 0. & 0. & 7.7 & 0. & 0. \\ 1.5 & 3.0 & 0. & 0. & 5.7 & 8.1 & 0. & 0. \\ 1.5 & 3.0 & 0. & 6.9 & 0. & 0. & 0. & 10.8 \\ 1.5 & 0. & 3.7 & 0. & 0. & 0. & 4.0 & 10.6 \\ 1.5 & 3.0 & 0. & 0. & 5.7 & 0. & 5.0 & 10.8 \end{bmatrix} \qquad (23a)$$

$$\mathbf{C} \cdot \mathbf{Tr} = \begin{bmatrix} 1.5 & 0. & 3.7 & 0. & 0. & 7.7 & 0. & 0. \\ 1.5 & 0. & 3.2 & 0. & 0. & 8.1 & 3.8 & 9.5 \\ 1.5 & 0. & 3.2 & 6.9 & 4.1 & 0. & 3.8 & 0. \\ 1.5 & 2.5 & 0. & 5.7 & 0. & 0. & 4.0 & 0. \\ 1.5 & 0. & 3.2 & 0. & 0. & 7.6 & 0. & 10.8 \end{bmatrix} \qquad (23b)$$

The vector **Sc** (24a and 24b) is obtained by summing the rows of the matrix **C•Tr**, so its coordinates determine the test values of the possible trajectories:

$$\mathbf{Sc} = \begin{bmatrix} 12.9 \\ 18.3 \\ 22.2 \\ 19.8 \\ 26.0 \end{bmatrix} \qquad (24a) \qquad\qquad \mathbf{Sc} = \begin{bmatrix} 12.9 \\ 26.1 \\ 19.5 \\ 13.7 \\ 23.1 \end{bmatrix} \qquad (24b)$$

The vectors **Sc** (24a and 24b) contain the values of each test trajectory according to the complex tests.

Conclusions

In the course of this research, we used the transition probability matrix prescribed for the connection type tracks A and B. These are summarized of the following two tables:

Table 1
The elements of the transition probability matrix $\left[\alpha_{i,j}\right]$ in *Fig. 1a* are determined by the distribution values of the large-scale network model

i/j	1	2	3	4	5	6	7	8
1								
2	$\alpha_{21} = P(2\|1)$							
3	$\alpha_{31} = P(3\|1)$							
4		$\alpha_{42} = P(4\|2)$						
5		$\alpha_{52} = P(5\|2)$						
6			$\alpha_{63} = P(6\|3)$		$\alpha_{65} = P(6\|5)$			
7			$\alpha_{73} = P(7\|3)$		$\alpha_{75} = P(7\|5)$			
8				$\alpha_{84} = P(8\|4)$			$\alpha_{87} = P(8\|7)$	

Table 2
The elements of the transition probability matrix $\left[\alpha_{i,j}\right]$ in *Fig. 1b* are determined by the distribution values of the large-scale network model

i/j	1	2	3	4	5	6	7	8
1								
2	$\alpha_{21} = P(2\|1)$							
3	$\alpha_{31} = P(3\|1)$							
4		$\alpha_{42} = P(4\|2)$			$\alpha_{45} = P(4\|5)$			
5			$\alpha_{53} = P(5\|3)$					
6			$\alpha_{63} = P(6\|3)$				$\alpha_{67} = P(6\|7)$	
7				$\alpha_{74} = P(7\|4)$				
8						$\alpha_{86} = P(8\|6)$		

The aim of our research was to develop an efficient and fast algorithm, that resulted in the determination of all the different trajectories, starting from any sector of the test track and ending in one or more sectors of an arbitrary choice.

In our opinion, in the case of interconnected curves the subject of the analysis is not only the value of the individual curves (in addition to the value of the traffic situations generated on the track and the value due to the geometric properties of the routing), but also, the value of which type of curve is connected to which other type of curve and how it impacts the driving, i.e. the value of the transition type.

For this purpose, our research used the connection matrix of the large-scale road network model.

The work is used for the qualification, development and design of test tracks, as it can also be used to perform complex evaluations of the defined trajectories. The calculations consider both the static and dynamic state characteristics of the sectors. They also take into account the geometric values of the sectors, the values of the situations generated on them, the values of the connections of the sectors and the direction of traversal of the sectors, which can also be freely varied. The field of application of the research is the definition and selection of the most valuable test trajectories, in summary, it helps the development of the capabilities of the test track and supports its successful operation. In the case of the examined trajectories, knowing the maximum number of sectors "m", the vector containing the routes can be determined in one step using the m-th power of the transition probability matrix. By analyzing the distributivity markers with a character analysis program, the trajectories can be automatically determined as described. The presented method accurately discusses and examines the geometry of the trajectories in an exact way. The presented matrix transformation expansion method for generating trajectories is extremely fast and very efficient in terms of real-time computational resources, which is a great advantage over the Xcity method that requires extraordinary computational power. The discussed method for evaluating test tracks already built or under design can be effectively applied both for the purposes of testing and for the selection of the most valuable trajectories defined by the objective function. The method can be similarly applied to the selection of optimal solutions for the exploration and evaluation of the possibilities of not yet constructed but pre-designed tracks. This provides an important opportunity for the preliminary evaluation of a large number of tracks, designed and prepared using computers, in the design phases and for further planning as needed. The above procedure takes advantage of a very useful property of the macroscopic model [8-13], that for, any complex transport network, the mathematical model, can take into account, all of the connections between the sector elements that make up the trajectories.

Acknowledgement

The research presented in this paper was carried out as part of the "Autonomous Vehicle Systems Research related to the Autonomous Vehicle Proving Ground of Zalaegerszeg (EFOP-3.6.2-16-2017-00002)" project in the framework of the New Széchenyi Plan. The completion of this project is funded by the European Union and co-financed by the European Social Fund.

References

[1] ACM. THE PROJECT (2018) http://www.acmwillowrun.org/the-project/

[2] R. Chen, M. Arief, and D. Zhao (2018) An "Xcity" Optimization Approach to Designing Proving Grounds for Connected and Autonomous Vehicles

Submission Date: August 1, 2018, pp. 1-18, https://arxiv.org/pdf/1808.03089.pdf

[3] Harman (2017) To Co-Develop and Operate the International Cyber Security Smart Mobility Analysis and Research Test (SMART) Range in Israel, 2017, url: https://news.harman.com/releases/releases-20171114

[4] Mcity Test Facility (2018) https://mcity.umich.edu/our-work/mcity-test-facility/

[5] D. Muoio (2018) Uber built a fake city in Pittsburgh with roaming mannequins to test its selfdriving cars. 2018, https://www.businessinsider.com/ubers-fake-city-pittsburgh-self-driving-cars-2017-10

[6] C. Nowakowski, S. E. Shladover, C.-Y. Chan, and H.-S. Tan. (2015) "Development of California Regulations to Govern Testing and Operation of Automated Driving Systems". Transportation Research Record: Journal of the Transportation Research Board 2489 (2015) pp. 137-144, doi: 10.3141/2489- 16. eprint: https://doi.org/10.3141/2489- 16, url: https://doi.org/10.3141/2489-16

[7] I. Passchier, G. v. Vugt, and M. Tideman (2015) "An Integral Approach to Autonomous and Cooperative, Vehicles Development and Testing". 2015 IEEE 18th International Conference on Intelligent Transportation Systems. Sept. 2015, pp. 348-352, doi: 10.1109/ITSC.2015. 66

[8] T. Péter and J. Bokor (2010) Research for the modelling and control of traffic, FISITA World Automotive Congress, Budapest, 30 May - 4 June 2010. Book of abstracts, pp. 66-73, In: Scientific Society for Mechanical Engineering (ISBN: 978-963-9058-28-6)

[9] T. Péter and J. Bokor (2011) New road traffic networks models for control, *GSTF International Journal on Computing*, Vol. 1, No. 2, pp. 227-232, DOI: 10.5176_2010-2283_1.2.65 February 2011

[10] T. Péter (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

[11] T. Péter and K. Szabó (2012) A new network model for the analysis of air traffic networks. In: Peridoica Polytechnica-Transportation Engineering 40/1 (2012) 39-44, doi: 10.3311/pp.tr.2012-1.07, web: http://www.pp.bme.hu/ tr ISSN 1587-3811 (online version); ISSN 0303-7800 (paper version)

[12] T. Peter, J. Bokor and A. Strobl (2013) Model for the analysis of traffic networks and traffic modelling of Győr, pp. 167-172, Doi: 0023, IFAC Workshop on Advances in Control and Automation Theory for Transportation Applications (ACATTA 2013) which is to be held in Istanbul, Turkey, 16-17 September 2013, http://www.acatta13.itu.edu.tr/

[13] T. Péter and S. Fazekas (2014) Determination of vehicle density of inputs and outputs and model validation for the analysis of network traffic processes, Periodica Polytechnica, Transportation Engineering Vol. 42, No 1, 2014, pp. 53-61

[14] T. Stevens (2017) Crashing Castle: An autonomous ride in Waymo's playground. 2017, url: https://www.cnet.com/roadshow/news/google-waymo-castle-visit/

[15] Z. Szalay (2016) "Structure and Architecture Problems of Autonomous Road Vehicle Testing and Validation". Structure and Architecture Problems of Autonomous Road Vehicle Testing and Validation. An optional note. 15th Mini Conference on Vehicle System Dynamics. The publisher, Nov. 2016

[16] Torc (2018) Asimov self-driving car system capabilities. 2018, url: https://torc.ai/wp-content/uploads/Torc_capabilities

[17] ZalaZone url: https://zalazone.hu/en/

[18] Kovács,T., Bolla, K., Gil, R.,A., Csizmás, E., Fábián, Cs., Kovács, L., Medgyes, K., Osztényi, J., Végh A. (2016) Parameters of the intelligent driver model in signalized intersections *Technical gazette*, Vol. 23, No. 5, October 2016, pp. 1469-1474, ISSN 1330-3651 (Print), ISSN 1848-6339 (Online) DOI: 10.17559/TV-20140702174255

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

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

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

[22] 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

[23] 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

[24] 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)]

Software Defect Prediction using Deep Learning

Meetesh Nevendra* **and Pradeep Singh**

Department of Computer Science and Engineering
National Institute of Technology, Raipur, India
e-mail: mnevendra.phd2018.cs@nitrr.ac.in and psingh.cs@nitrr.ac.in

Abstract: An increasing number of defects in software, damages the quality and reliability of that software. The detection of defective instances is becoming increasingly important, and current detection techniques require a great deal of improvement. However, Machine Learning (ML) techniques are effectively used, to detect defects in software. The primary purpose of ML techniques in Software Defect Prediction (SDP) is to predict defects, according to historical data. Establishing a critical SDP model on high-dimensional and limited data is still a challenging task. Thus, in this paper, we proposed an approach to detect defective modules in software using enhanced Convolutional Neural Networks (CNNs). The paper aims to identify the defective instance using the enhanced deep learning method. Our experiments are based on Within Project Defect Prediction (WPDP), where K-fold cross-validation is performed. The proposed approach has been evaluated on nineteen open-source software defect datasets, with respect to different evaluation metrics. Empirical results show that our proposed approach is significantly better than Li's CNN and standard ML model. In addition, we performed the Scott-Knot ESD test, which shows the effectiveness of our proposed approach.

Keywords: Software defect; CNN; Deep learning

1 Introduction

In the modern area, software quality is increasing rapidly, which directly affects the cost and reliability of the software product. However, the presence of defects in the software linearly decreases the quality and increases the software product's cost. SDP [1-3] is a technique that builds a classifier and predicts the code areas that potentially contain defects. The classifier's outcomes (i.e., defective programming regions) can locate indications for code reviewers to allocate their efforts. SDP is an essential part of software quality analysis [4] and is analyzed through software reliability engineering. In software engineering, early detection of defective parts of a software system can help developers and engineers in finding the correct way to use the limited resources in the testing and maintenance phases of software development.

Several research studies have been done for SDP to predict defective instances from historical data [1] [5]. SDP is more feasible because it can predict the defective instances and ensure developers from where defects arise. In recent work, Dam et al. proposed experimental research on SDP using a deep tree-based prediction model [6].

Nowadays, deep learning (DL) has played an essential role in the ML literature [7], and it has been used by different research areas and proven to be very useful, especially in speech recognition [8] and image processing [9]. Still, the use of DL for SDP is not thoroughly investigated. This study develops an enhanced deep learning model to investigate how the deep learning model will work with defect prediction datasets. We proposed an approach that utilizes enhanced CNN to predict the defective instances from a historical dataset. The proposed approach comprises two stages: model construction and prediction. In the model-construction phase, we first select the appropriate features using the feature selection (FS) technique, then these features are used to build the model using an enhanced CNN approach. Once the model is built, we predict the software defectiveness in the prediction phase.

To evaluate our approach, we used accuracy, precision, recall, and f1-score, one of the most widely used metrics. We conducted experiments on 19 open source software defect datasets. The experimental results suggest that the suggested approach performs better than Li's CNN and standard ML models. We also performed the Scott-Knot ESD (SK-ESD) test to indicate that our proposed approach has utility.

The remaining of our paper is summarized as follows. In Section 2, we discuss related works. Section 3 provides an overview and description of the CNN architecture. Section 4 the datasets and performance measures are given. Section 5 shows the overall outline of our approach. Section 6 describes our experiments and results. Finally, Section 7 offers Conclusions and future work.

2 Related Work

In order to more understand the SDP, we look back at the previous work done by the researchers in past years. Before going into SDP for details, we do an in-depth review of valuable methods that are indifferent to software metrics bases and well-known in this area of SDP. Chidamber et al. [10] explain that software metrics are suited for SDP. Basili et al. [11] examined and validated the object-oriented design metrics (OODM) and determines that either this OODM is valid or not for SDP.

The reliability of the software varies according to different coding styles and various other parameters. Reliability is also one of the critical issues in SDP. To solve the problem, additional knowledge needs to be gathered in the form of historical source code. However, this problem is addressed by Gyimothy et al. [12] and by a new method based on object-oriented metrics analysis and determined that these metrics are highly suitable for improving the model's prediction performance. Singh and Verma [13] also performed the defect prediction, in the design phase. They explained that design metrics are beneficial for the software development life cycle (SDLC), and the prediction of defects should happen at the initial phase of SDLC for early warning and cost-effective software development.

Several prior approaches have been investigated for classifying the SDP. Byoung et al. [14] established a novel polynomial function-based neural network (pf-NN) model for SDP. The approach aggregates fuzzy C-means and genetic clustering techniques, facilitating the acquisition of nonlinear parameters for a procedure. However, in terms of classification, clustering techniques are not suitable for achieving enhanced prediction performance. Different ML algorithms have been used for SDP to overcome this problem, such as Naïve Bayes (NB) classifier, support vector machine (SVM), decision tree, neural network, and DL techniques. Elish et al. [15] established the SDP scheme using SVM. They evaluated the effectiveness of SVM against eight statistical and ML models in the context of four open-source NASA datasets. The outcomes demonstrate that SVM is more effective in finding defects compare to other models. Researchers also found that NB is suitable for classifying SDP. Shivaji et al. [16] employed the NB classification techniques for SDP using FS methods. They found that NB using FS improves the significant performance of defective f-measure by 21%. However, they also show that NB achieves a 12% improvement over SVM. Correspondingly, Dejaeger et al. [17] performs the SDP using 15 different Bayesian networks and other popular ML methods and found that augmented NB classifiers perform better than other classifiers in terms of the ROC curve.

Santosh et al. [18] also performed the SDP using decision tree classification and developed a recommendation system for SDP. They found that the tree classification technique is more suitable for finding the defects. Singh and Verma [19] utilized 16 open-source datasets to find the defects using a multi classifier approach, a combination of SVM, NB, and Random forest (RF). They found that this technique is more effective for performing the SDP. Singh et al. [20] utilized the fuzzy rule-based approach for finding the defect in software metrics. They find that a fuzzy rule-based classification technique can produce competitive or improved performance than C4.5, RF, and NB classifiers. They also determine that the proposed technique is a more comprehensive option than the other existing techniques for understanding several aspects that determine software defects.

Recently, Yang et al. [21] utilized a DL-based technique for SDP. They utilized a deep belief network to predict defects on six large open-source projects. The experimental results show that the proposed approach can achieve significant results than other approaches. Manjula and Florence [22] also developed a deep based hybrid approach for SDP using software metrics. They combine the genetic algorithm and deep neural network for classification. The outcomes showed that the proposed approach performs significantly better than other techniques. Li et al. [23]

and Phan et al. [24] utilized CNN to predict the software defect. Also, Zhao et al. [25] predict the software defect via cost-sensitive siamese parallel fully connected neural networks. The outcomes of these studies show that the DL technique plays an essential role in ML for SDP and can be utilized in several classification areas.

However, these techniques still go through many problems such as accuracy, computational time, and complexity with respect to defect prediction. Nevertheless, these problems can be further improved. Here we present an enhanced CNN approach that helps to reduce the overfitting problem and provide a significantly better classification rate to overcome these issues. The entire process of our proposed model is shown in Section 5.

3 Convolutional Neural Network Architecture

Our proposed approach enhances the CNN model to predict software defect instances in software defect datasets. Our CNN model has four convolution layers, two pooling layers, a flattening layer, and two dense layers. Li's CNN model and our enhanced CNN model are compared in Table 1.

Table 1

Li's CNN model compared to our enhanced CNN model

	Li’s CNN	Enhanced CNN
Convolutional Layers	One	Four
Embedding Layer	✓	×
Dense layer	One	Two
Pooling Layers	One	Two
Dropout	×	✓
Activation function	ReLU and sigmoid	ReLU and sigmoid
Training and optimizer	Mini-batch SGD and Adam	Adam and binary cross-entropy
Parameter initialization	×	✓

Our enhanced CNN model and Li's model have several changes, like convolutional layers, pooling layers, dense layers and dropout layers, activation function, optimizer and parameter initializer. We utilized one dropout between convolutional layers and one dropout between dense layers in our enhanced CNN architecture. These changes in the model will help to increase the performance and reduce the problem of overfitting.

3.1 Convolution Layer

CNN relies heavily on convolutional layers as a building block. In a convolutional layer, the goal is to extract features from the input data. Each layer has a set of learnable filters as its parameter $w = w_1, w_2, \dots, w_n$ and biases = $b = b_1, b_2, \dots, b_n$. Layers apply convolution operations to generate a feature map X_n and pass the result on to subsequent layers. A nonlinear element-wise transform $\sigma(\cdot)$ is applied to these features, and the same process is repeated for each convolutional layer k.

$$X_n^k = \sigma(w_n^{k-1} * X^{k-1} + b_n^{k-1}) \tag{1}$$

3.2 Pooling Layer

Using a pooling layer, which reduces the resolution of the feature maps to achieve shift-invariance, is usually placed between two convolutional layers. Each feature map in a pooling layer is connected to its corresponding feature map in the convolutional layer preceding it in the pipeline. For each feature map $a_{:,:,k}^l$ we have $pool(\cdot)$ as the pooling function.

$$y_{i,j,k}^l = pool\big(a_{m,n,k}^l\big), \forall(m,n) \in \mathcal{R}_{i,j} \tag{2}$$

where $\mathcal{R}_{i,j}$ is a local neighbourhood around location (i,j). Average pooling and maximum pooling are the two most common pooling operations [26].

3.3 Flatten Layer

It converts the data into a 1-dimensional array so that it can be input into the next layer. We flatten the output of the convolutional layers to create a single long feature vector. A fully connected layer links it to the final classification model.

3.4 Dropout

Dropout was first introduced by Hinton et al. [27], and it has been shown to be very effective in reducing overfitting. As part of our enhanced CNN architecture, we apply dropout after the max-pooling and the fully connected layers, respectively. Dropout has the following output:

$$y = r * a(W^T x) \tag{3}$$

where $x = [x_1, x_2, \dots, x_n]^T$ is the input to a layer that is fully connected $W^T \in \mathcal{R}^{n*d}$ is a weight matrix and r is a binary vector of size d.

3.5 Dense Layer

The CNN has dense layers after convolution and pooling. It's important to note that each node in the dense layer is fully connected to every other node in previous layers. Dense layers integrate local information with category differentiation in the convolutional or pooling layer, which is the function of the layer. This equation can be represented as:

$$dl_1 = f(\sum_{p=1}^{N} \mathfrak{w}_{1,p} * o_p \ + b) \tag{4}$$

A neuron's activation function is denoted by the f, where $\mathfrak{w}$ is a weight vector, o is the input vector, and b is the bias value.

For this paper, the dense layer's activation function was referred to as Rectified Linear Unit (ReLU). The ReLU is treated as a standard activation function in CNN, however. They have shown that they are faster to train than standard sigmoid units in the hidden layer and can sometimes help with discriminative performance [28]. In this case, the derivative of the ReLU activation function is given as:

$$f'(x) = \frac{\delta f(x)}{\delta x} = \begin{cases} 0 \ if \ x < 0 \\ 1 \ if \ x > 0 \end{cases} \tag{5}$$

Our CNN architecture utilized the sigmoid activation function (SAF) at the last layer (output layer). The SAF is used to find a probability between 0 to 1. Mainly SAF is utilized where the probability needs to be found as an output. However, the probability has occurred only between 0 to 1; thus, the SAF is the correct choice. Mathematically the sigmoid activation function is defined as:

$$S(x) = \frac{1}{1+e^{-x}} \tag{6}$$

4 Datasets and Performance Measure

4.1 Datasets

To estimate the prediction capabilities of our CNN model, we experimented on 19 open-source software defect datasets. These defect datasets are collected from the tera-PROMISE data repository [29]. The instance of every dataset corresponds to two parts: metrics and labels. Table 2 show the statistics of utilized datasets. Columns one and four represent the dataset ID, columns two and six represent the dataset name, columns three and six represent the line of code, and columns four and eight represent the number of instances and defects. However, Table 3 shown the features present in the dataset. Columns 1 and 3 represent the feature ID, and columns 2 and 4 represent the features name of all the datasets.

Table 2

Statistics of Datasets

D.ID	Dataset	LOC	Instance / Defects	D.ID	Dataset	LOC	Instance / Defects
D01	log4j-1.0	21,549	135 / 34	D11	synapse-1.1	42,302	222 / 60
D02	log4j-1.2	38,191	205 / 189	D12	synapse-1.2	53,500	256 / 86
D03	lucene-2.0	50,596	195 / 91	D13	velocity-1.4	51,713	196 / 147
D04	lucene-2.2	63,571	247 / 144	D14	velocity-1.6	57,012	229 / 78

D05	lucene-2.4	102,859	340 / 203	D15	xalan-2.4	225,088	723 / 110
D06	poi-1.5	55,428	237 / 141	D16	xalan-2.5	304,860	803 / 387
D07	poi-2.0	93,171	314 / 37	D17	xalan-2.6	411,737	885 / 411
D08	poi-2.5	119,731	385 / 248	D18	xerces-1.2	159,254	440 / 71
D09	poi-3.0	129,327	442 / 281	D19	xerces-1.3	167,095	453 / 69
D10	synapse-1.0	159,254	440 / 71	-	Total	2,306,238	7,147 / 2,858

Table 3

Features in the datasets

F.ID	Features name	F.ID	Features name
1	wmc	11	moa
2	dit	12	mfa
3	noc	13	cam
4	cbo	14	ic
5	rfc	15	cbm
6	lcom	16	amc
7	lcom3	17	ca
8	npm	18	ce
9	loc	19	max_cc
10	dam	20	avg_cc

4.2 Performance Measure

In order to compare the results, we evaluate measures such as:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

The accuracy is the percentage of correct classifications in the total number of classifications. Where T and F stand for true and false, P and N stand for positive and negative samples, respectively.

$$Precision = \frac{TP}{TP+FP} \tag{8}$$

The precision is calculated by dividing the number of correct classifications by the number of incorrect.

$$Recall = \frac{TP}{TP+FN} \tag{9}$$

The recall measures the number of correct classifications minus the number of missed entries.

$$F1 - score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{10}$$

On the other hand, an F1-score is a derived effectiveness measurement that measures the harmonic mean of precision and recall.

5 Proposed Approach

The overall workflow of our proposed approach is shown in Figure 1 below. The proposed enhanced CNN model was applied for predicting the defects in software projects. There are two phases to this approach: model construction and prediction. For model construction, the dataset is divided into k-folds, where k-1 folds are used to train the CNN model, and one fold is used to test the model.

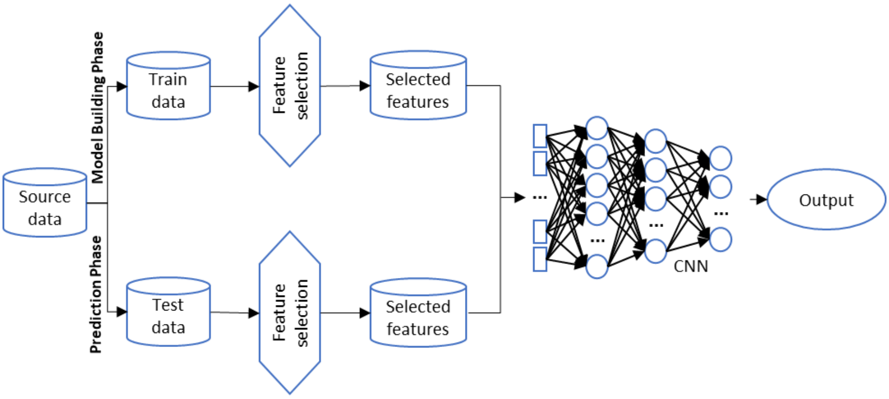


Figure 1
The overall workflow of our proposed approach

In order to train the CNN model, the FS technique is applied to the selected k-1 parts of data because the CNN model's input size should be expressed in metric units of $n*n$. The $n*n$ metrics cannot accommodate m number of features in our dataset. So, either increase the number of features in the dataset or remove those that aren't needed. This is why we removed unnecessary features to convert our dataset into $n*n$ format in order to avoid performance degradation. We used Chi-square (Chi2) filter-based FS to select N number of features in our experiment. It has been demonstrated by Nam et al. [30] that Chi-square is the most effective FS of all approaches. When features and classes are linked together, Chi2 performs well. When selecting relevant features, it helps to identify the frequency between class and feature. In our execution, we selected only those features that had the highest Chi2 scores. Calculation of the Chi2 score is denoted by:

$$Chi - square\ (\chi^2) = \frac{(observed\ frequency - Expected\ frequency)^2}{Expected\ frequency} \quad (11)$$

The number of examined classes is defined as the observed frequency. However, if there were no association between feature and target, the predicted class number would represent the expected frequency of that feature. Chi2 FS ranked the features based on their relationship to the class as long as they are ranked in order. To create n*n features metrics, we removed the lower-ranked feature metric from the features metrics. $n * n$ 2D metrics are created after N features are chosen from the source data.

Algorithm 1 Proposed Approach

```
Input:            $Dataset\ \mathcal{D} = \{x_t, y_t\}_{t=1}^{n}$
                  $Candidate\ CNN\ algorithm$
Output:           $Predict\ the\ performance\ score$
Initialization:   $Select\ k = 10\ for\ k - fold\ cross - validation$
                  $N = 10\ for\ number\ of\ runs$
Procedure:        $for\ i = 1\ to\ N$
                    $for\ each\ fold\ k\ do$
                      $Train, Test = \mathcal{D}(test\ size = 0.9, train\ size = 0.1)$
                      $Select\ m\ top\ features\ from\ Train\ data$
                      $train_{new} = \chi^2 = \sum_{r=1}^{p}\sum_{s=1}^{q}\frac{(D_{rs}-E_{rs})^2}{E_{rs}}$
                      $Model = CNN(train_{new})$
                      $test_{new} = Select\ same\ features\ from\ Test\ as\ train_{new}$
                      $Predict_{test} = Model.predict(test_{new})$
                      $Performance\ score\ [k] = Performance(Predict_{test})$
                    $end\ for$
                    $Performance\ score\ [i] = Performance[k].average$ ,
                  $end\ for$
                  $Final_{Performance} = Performance[i].average$
```

Further, the two-dimensional metric is transformed into 3D feature maps. Then the 3D metrics are then used to train a CNN model based on these metrics. In the prediction phase, the generated model is used to predict the defects (defective or non-defective) from the remaining one part of the data; before the prediction phase, the remaining one part of the data is selected as the same as features like the k-1 part of the data. The experiments are conducted ten times, and the average of all runs is selected as an outcome. The final outcomes are compared with Li's CNN and benchmarking ML approaches, SVM, AdaBoost, and KNN. Wu et al. [31] also added these three algorithms as the top 10 algorithms in data mining. Algorithm 1 shows the implementation of our proposed approach.

The SVM [32] is one of the effective and accurate supervised ML approaches. However, SVM has to find the best classification function to distinguish between members of two classes (0 and 1) in the training dataset.

The AdaBoost algorithm [33] was recommended by Freund and Schapire in 1977 and found one of the essential ensembles approaches since it has a substantial theoretical establishment, extremely accurate prediction, incredible simplicity, and extensive applications. The AdaBoost primary function is to generate a classifier by using the base learning algorithms repeatedly. The AdaBoost algorithm is used for both classification and regression problems.

K-nearest neighbor (KNN) [33] classifier has used a cluster of k substances in a training dataset neighboring the test entity. It bases the allocation of a label on the majority of a specific class in this neighborhood. The algorithm computes the distance of the z object from the test object with training objects to establish its nearest neighbor. Once the nearest-neighbor list is achieved, the test object is classified based on its nearest neighbors' majority class and classifies their respective classes.

6 Results

For the experiments, we utilized a 10-fold cross-validation technique [34] to evaluate the execution of the proposed model in within-version scenarios. The proposed approach divides the datasets into ten equal parts, each with equal features and classes. We need to create the $n * n$ metrics from the given feature metrics, and it can't be possible to convert the $n * n$ metrics from the given 20 feature metrics. So to transform our data into $n * n$ metrics, we choose the top 16 features out of 20 features. In order to transform metrics into $4 * 4$ 2D feature metrics, we select the 16 most important features from each metric. As shown in Table 4, the selected feature ID can be found. A total of 16 features are selected for each dataset that is specified. According to their class, the features are chosen. First, the feature with the highest relative importance is chosen, followed by the second, and so on. The features that have been selected are displayed in decreasing order. It's crucial to select the maximum relevance features first and then minor relevance features at the end.

Table 4

Selected features ID for proposed approach

D01	D02	D03	D04	D05	D06	D07	D08	D09	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19
11	6	11	6	11	6	11	11	11	11	11	11	11	11	11	6	11	11	11
6	18	6	11	6	11	6	6	6	5	6	5	18	6	6	11	6	6	6
5	11	5	5	5	5	18	5	5	19	5	6	6	5	5	5	18	5	5
4	7	18	1	4	1	5	1	1	18	8	4	5	18	18	1	5	18	18
7	4	1	9	1	9	1	9	9	6	19	18	19	8	1	9	1	1	8
1	5	19	4	9	18	8	17	18	8	1	8	1	1	9	19	9	13	1
9	9	8	8	7	8	4	8	4	1	18	19	17	9	19	18	4	8	4
19	1	4	7	8	3	19	7	8	7	4	1	16	4	4	17	19	17	13
8	17	9	17	18	4	9	19	19	4	9	7	7	19	7	4	7	4	19
18	3	3	18	17	17	13	16	17	9	13	9	3	3	8	7	8	9	17
20	8	13	13	3	13	17	3	13	3	7	17	4	13	17	3	13	3	7
13	12	7	10	19	19	7	13	7	20	12	13	2	17	13	13	3	7	16
17	16	17	3	13	16	3	2	16	17	14	16	8	15	20	8	17	19	9
16	19	12	19	16	2	20	14	20	16	20	12	20	10	3	20	10	2	12
12	14	16	12	12	14	16	10	12	13	2	20	14	16	16	16	20	14	3
3	20	8	15	10	7	14	4	10	15	10	10	12	2	15	2	14	20	10

After converting 4*4 2D metrics to 3D metrics, feed this 3D metrics data into our enhanced CNN model. Before the prediction phase, the test data feature metrics are converted into 3D metrics so that they are the same as the source data.

Every fold of the data is used ten times in the execution of the program. In every iteration, the model's training is performed with nine parts of the data; however, the remaining part is used for testing. For the statistical reliability of the result, we run our experiment 10 times and record the average performance. However, we used accuracy, precision, recall and f1-score as a performance measure.

Table 5 shows the obtained result of our proposed approach vs different ML techniques, the best outcome of the average result in bold. We found that the proposed approach performs significantly better than different "state-of-the-art" ML approaches. The proposed model shows improvement with respect to all the performance evaluations. Proposed model overcome the KNN, SVM and AdaBoost with respect to accuracy by 3.68%, 5.98%, 2.48%, with respect to precision by 3.51%, 5.79%, 2.07%, with respect to recall by 2.95%, 4.8%, 3.46% and with respect to f1-score by 3.23%, 5.35% and 2.45% respectively. However, to identify the significance of our proposed model, we applied the SK-ESD test, which is implemented by Tantithamthavorn et al. [35]. It is also available on CRAN[1].

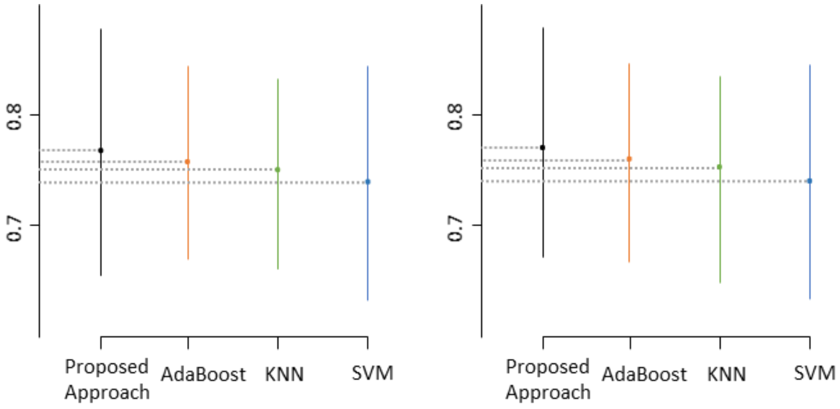
Table 5
Comparison of Proposed Approach

D06	D07	D08	D09	D10	D11	D12	D13	D14	D15	D16	D17	D18	D19	Average
0.696	0.875	0.761	0.751	0.872	0.716	0.73	0.729	0.654	0.842	0.671	0.714	0.806	0.823	0.747
0.726	0.853	0.75	0.762	0.815	0.723	0.735	0.751	0.674	0.842	0.643	0.697	0.847	0.836	0.755
0.749	0.825	0.782	0.78	0.815	0.719	0.734	0.792	0.687	0.875	0.663	0.7	0.853	0.875	0.767
0.737	0.839	0.766	0.771	0.815	0.721	0.734	0.771	0.68	0.858	0.653	0.698	0.85	0.855	0.761
0.674	0.882	0.706	0.742	0.897	0.734	0.672	0.759	0.676	0.846	0.596	0.618	0.834	0.852	0.73
0.705	0.854	0.732	0.759	0.892	0.732	0.641	0.753	0.685	0.856	0.589	0.658	0.824	0.846	0.738
0.721	0.861	0.744	0.791	0.871	0.724	0.674	0.753	0.679	0.84	0.611	0.695	0.854	0.883	0.753
0.713	0.857	0.738	0.775	0.881	0.728	0.657	0.753	0.682	0.848	0.6	0.676	0.839	0.864	0.745
0.717	0.841	0.802	0.778	0.853	0.765	0.718	0.841	0.703	0.817	0.651	0.733	0.779	0.834	0.756
0.771	0.824	0.871	0.832	0.831	0.739	0.602	0.858	0.775	0.799	0.628	0.674	0.753	0.822	0.766
0.754	0.864	0.837	0.833	0.833	0.711	0.609	0.87	0.738	0.829	0.643	0.734	0.755	0.852	0.769
0.762	0.844	0.854	0.832	0.832	0.725	0.605	0.864	0.756	0.814	0.635	0.703	0.754	0.837	0.767
0.696	0.883	0.744	0.898	0.897	0.729	0.749	0.694	0.847	0.848	0.586	0.839	0.847	0.842	**0.775**
0.729	0.871	0.753	0.902	0.883	0.734	0.765	0.704	0.834	0.821	0.561	0.862	0.876	0.834	**0.782**
0.732	0.886	0.76	0.91	0.886	0.732	0.754	0.71	0.859	0.829	0.584	0.856	0.888	0.856	**0.79**
0.73	0.878	0.756	0.906	0.884	0.733	0.759	0.707	0.846	0.825	0.572	0.859	0.882	0.845	**0.786**

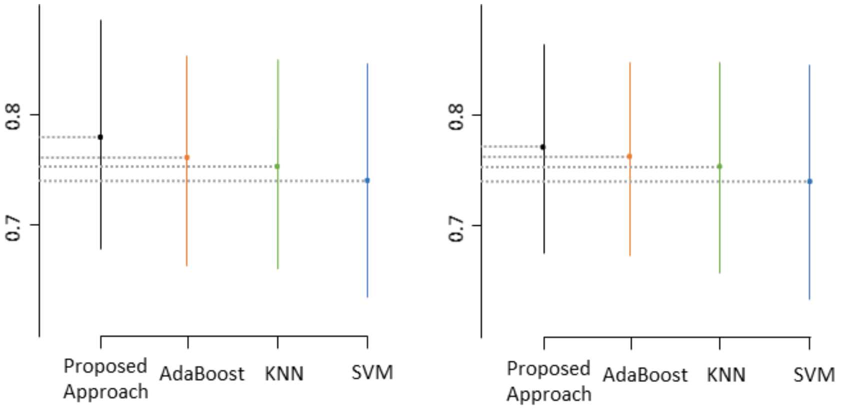
[1] https://cran.r-project.org/web/packages/ScottKnottESD/index.html

DATA		D01	D02	D03	D04	D05
KNN	Acc	0.731	0.917	0.641	0.63	0.629
	pre	0.752	0.92	0.54	0.714	0.761
	rec	0.783	0.926	0.645	0.667	0.707
	F1	0.767	0.923	0.588	0.69	0.733
SVM	Acc	0.739	0.921	0.471	0.607	0.644
	pre	0.724	0.912	0.545	0.653	0.654
	rec	0.753	0.922	0.566	0.659	0.697
	F1	0.738	0.917	0.555	0.656	0.675
AdaBoost	Acc	0.717	0.907	0.625	0.619	0.664
	Pre	0.775	0.958	0.631	0.667	0.741
	rec	0.752	0.943	0.656	0.645	0.747
	F1	0.763	0.95	0.643	0.656	0.744
Proposed Approach	Acc	0.748	0.942	0.628	0.618	0.697
	Pre	0.756	0.952	0.638	0.658	0.721
	rec	0.778	0.962	0.646	0.648	0.732
	F1	0.767	0.957	0.642	0.653	0.726

The SK-ESD test is a mean comparison approach. It is an alternative approach of the Scott-Knott test [36] that finds the magnitude difference of each means within a group and between groups. The SK-ESD test finds the mean ranking. We apply the SK-ESD test in order to find the magnitude difference between the proposed approach and other presented ML approaches.



(a) Accuracy (b) Precision



(c) Recall

(d) F1-score

Figure 2

Scott knot test result

Figure 2 shows the SK-ESD mean ranking comparison of the proposed approach. The result of the SK-ESD test lies between the proposed approach and present machine learning approaches. Figures 2(a), 2(b), 2(c) and 2(d) show the SK-ESD mean ranking comparison of accuracy, precision, recall and f1-score, respectively. In these figures, the horizontal grey dashed line indicates the mean value of all the presented methods, helping to visualize the mean differences between each method. We found that the proposed approach obtained the highest mean rank in all four scenarios from the figure. This shows that the proposed approach is performing better than the compared models for SDP. Table 6 also shows the obtained result of our proposed approach vs Li's CNN architecture. The best-performed result is in boldface. From Table 6, we can see that our proposed model is performed better than Li's CNN architecture. The proposed approach improves the performance over 15 datasets concerning all the applied performance evaluations. The proposed model overcomes Li's CNN model concerning the accuracy, precision, recall and f1-score by 15.12%, 16.32%, 16.3% and 16.52%, respectively.

Table 6

Performance comparison of enhancing CNN model with Li's CNN architecture

Data	Li's CNN				Enhanced CNN			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
D01	0.682	0.623	0.533	0.574	**0.748**	**0.756**	**0.778**	**0.767**
D02	0.922	0.902	0.910	0.906	**0.942**	**0.952**	**0.962**	**0.957**
D03	0.468	0.528	0.570	0.548	**0.628**	**0.638**	**0.646**	**0.642**
D04	0.583	0.573	0.568	0.570	**0.618**	**0.658**	**0.648**	**0.653**
D05	0.597	0.586	0.592	0.589	**0.697**	**0.721**	**0.732**	**0.726**
D06	0.596	0.573	0.563	0.568	**0.696**	**0.729**	**0.732**	**0.730**
D07	**0.883**	**0.872**	**0.888**	**0.880**	**0.883**	0.871	0.886	0.878

D08	0.644	0.653	0.672	0.662	**0.744**	**0.753**	**0.76**	**0.756**
D09	0.636	0.643	0.657	0.650	**0.898**	**0.902**	**0.91**	**0.906**
D10	**0.898**	**0.900**	**0.905**	**0.902**	0.897	0.883	0.886	0.884
D11	0.720	0.718	**0.738**	0.728	**0.729**	**0.734**	0.732	**0.733**
D12	0.462	0.462	0.467	0.464	**0.749**	**0.765**	**0.754**	**0.759**
D13	**0.728**	**0.792**	**0.870**	**0.829**	0.694	0.704	0.71	0.707
D14	0.371	0.345	0.381	0.362	**0.847**	**0.834**	**0.859**	**0.846**
D15	**0.849**	**0.856**	**0.832**	**0.844**	0.848	0.821	0.829	0.825
D16	0.491	0.486	0.489	0.487	**0.586**	**0.561**	**0.584**	**0.572**
D17	0.470	0.467	0.476	0.471	**0.839**	**0.862**	**0.856**	**0.859**
D18	0.839	0.829	0.812	0.820	**0.847**	**0.876**	**0.888**	**0.882**
D19	0.822	0.810	0.801	0.805	**0.842**	**0.834**	**0.856**	**0.845**
Average	0.666	0.664	0.670	0.666	**0.775**	**0.782**	**0.790**	**0.786**

In conclusion, we can say that the performance of enhanced CNN depends on the input data and model architecture. The data which are less valuable are eliminated to train the model. However, carefully increasing the architecture and parameters of the CNN model will lead to enhance the model. As a result, it enhances the training process and become a good prediction model. Compared to all other methods, the proposed CNN model performed significantly better.

Conclusions

The early detection and prediction of software defects plays an important role in modern software development, in terms of effective resource allocation. To address this issue of SDP, in this paper, we developed an enhanced CNN approach. First, in this approach, FS is applied to select $n * n$ 2D metric in the training dataset, and further, this metric is mapped into the 3D metrics. The CNN model is then trained using the generated metrics; it can predict the defective instances once the model is trained. The proposed enhanced CNN model significantly improves compared to existing benchmark classification schemes such as KNN, SVM, AdaBoost and Li's CNN model. We performed 10-fold cross-validation and calculated the average of all runs, resulting in the final result. Scott Knott ESD's mean ranking comparison of the proposed approach and other presented ML approaches shows that the proposed approach achieves the highest ranking.

Future research will focus on time reduction and accelerated network training. Also, exploration of other software metrics, aimed at the development of more efficient DL models, will be considered.

References

[1] Arar, Ö. F., Ayan, K.: Software defect prediction using cost-sensitive neural network. *Appl. Soft Comput.*, 33, 2015, pp. 263-277

[2] Chen, X. et al.: Software defect number prediction: Unsupervised vs supervised methods. *Inf. Softw. Technol.*, 106, 2019, pp. 161-181

[3] Nevendra, M., Singh, P.: *Multistage Preprocessing Approach for Software Defect Data Prediction*. In: Communications in Computer and Information Science. 2018, pp. 505-515

[4] Miholca, D. L. et al.: A novel approach for software defect prediction through hybridizing gradual relational association rules with artificial neural networks. *Inf. Sci. (Ny).*, 441, 2018, pp. 152-170

[5] Akmel, F. et al.: A Literature Review Study of Software Defect Prediction using Machine Learning Techniques. *Int. J. Emerg. Res. Manag. Technol.*, 6 (6), 2018, p. 300

[6] Dam, H. K. et al.: A deep tree-based model for software defect prediction. *arXiv Prepr. arXiv1802.00921*, 2018

[7] Jordan, M. I., Mitchell, T. M.: Machine learning: Trends, perspectives, and prospects. *Science (80-.).*, 349 (6245), 2015, pp. 255-260

[8] Graves, A. et al.: Speech recognition with deep recurrent neural networks. *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, (6), 2013, pp. 6645-6649

[9] Affonso, C. et al.: Deep learning for biological image classification. *Expert Syst. Appl.*, 85, 2017, pp. 114-122

[10] Chidamber, S. R., Kemerer, C. F.: A Metrics Suite for Object Oriented Design. *IEEE Trans. Softw. Eng.*, 20 (6), 1994, pp. 476-493

[11] Basili, V. R. et al.: A validation of object-oriented design metrics as qualityindicators. *IEEE Trans. Softw. Eng.*, 22 (10), 1996, pp. 751-761

[12] Gyimothy, T. et al.: Empirical Validation of Object-Oriented Metrics on Open Source Software for Fault Prediction. *IEEE Trans. Softw. Eng.*, 31 (10), 2005, pp. 897-910

[13] Singh, P., Verma, S.: Cross Project Software Fault Prediction at Design Phase. *Int. J. Comput. Inf. Eng.*, 9 (3), 2015, pp. 800-805

[14] Park, B. J. et al.: The design of polynomial function-based neural network predictors for detection of software defects. *Inf. Sci. (Ny).*, 229, 2013, pp. 40-57

[15] Elish, K. O., Elish, M. O.: Predicting defect-prone software modules using support vector machines. *J. Syst. Softw.*, 81 (5), 2008, pp. 649-660

[16] Shivaji, S. et al.: Reducing features to improve code change-based bug prediction. *IEEE Trans. Softw. Eng.*, 39 (4), 2013, pp. 552-569

[17] Dejaeger, K. et al.: Toward comprehensible software fault prediction models using bayesian network classifiers. *IEEE Trans. Softw. Eng.*, 39 (2), 2013, pp. 237-257

[18] Rathore, S. S., Kumar, S.: A decision tree logic based recommendation

system to select software fault prediction techniques. *Computing*, 99 (3), 2016, pp. 1-31

[19] Singh, P., Verma, S.: Multi-classifier model for software fault prediction. *Int. Arab J. Inf. Technol.*, 15 (5), 2018, pp. 912-919

[20] Singh, P. et al.: Fuzzy Rule-Based Approach for Software Fault Prediction. *IEEE Trans. Syst. Man, Cybern. Syst.*, 47 (5), 2017, pp. 826-837

[21] Yang, X. et al.: *Deep Learning for Just-in-Time Defect Prediction*. In: 2015 IEEE International Conference on Software Quality, Reliability and Security. IEEE, 2015, pp. 17-26

[22] Manjula, C., Florence, L.: Deep neural network based hybrid approach for software defect prediction using software metrics. *Cluster Comput.*, 2018, pp. 1-17

[23] Li, J. et al.: *Software Defect Prediction via Convolutional Neural Network*. In: 2017 IEEE International Conference on Software Quality, Reliability and Security (QRS) IEEE, 2017, pp. 318-328

[24] Viet Phan, A. et al.: *Convolutional Neural Networks over Control Flow Graphs for Software Defect Prediction*. In: 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI) IEEE, 2017, pp. 45-52

[25] Zhao, L. et al.: Software defect prediction via cost-sensitive Siamese parallel fully-connected neural networks. *Neurocomputing*, 352, 2019, pp. 64-74

[26] Nagi, J. et al.: *Max-pooling convolutional neural networks for vision-based hand gesture recognition*. In: 2011 IEEE International Conference on Signal and Image Processing Applications (ICSIPA) IEEE, 2011, pp. 342-347

[27] Hinton, G. E. et al.: Improving neural networks by preventing co-adaptation of feature detectors. *arXiv Prepr. arXiv1207.0580*, 2012

[28] Dahl, G. E. et al.: *Improving deep neural networks for LVCSR using rectified linear units and dropout*. In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, 2013, pp. 8609-8613

[29] *tera-PROMISE: Welcome to one of the largest repositories of SE research data*. no date

[30] Nam, J., Kim, S.: *Heterogeneous defect prediction*. In: Proceedings of the 2015 10th Joint Meeting on Foundations of Software Engineering - ESEC/FSE 2015. New York, New York, USA: ACM Press, 2015, pp. 508-519

[31] Wu, X. et al.: Top 10 algorithms in data mining. *Knowl. Inf. Syst.*, 14 (1), 2008, pp. 1-37

[32] Vapnik, V.: *The nature of statistical learning theory.* Springer science & business media, 2013

[33] Fix, E.: *Discriminatory analysis: nonparametric discrimination, consistency properties*. USAF School of Aviation Medicine, 1951

[34] Kohavi, R.: A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *Int. Jt. Conf. Artif. Intell.*, 14 (2), 1995, pp. 1137-1145

[35] Tantithamthavorn, C. et al.: The Impact of Automated Parameter Optimization on Defect Prediction Models. *IEEE Trans. Softw. Eng.*, 45 (7), 2019, pp. 683-711

[36] Jelihovschi, E. G. et al.: The ScottKnott clustering algorithm. *Univ. Estadual St. Cruz-UESC, Ilheus, Bahia, Bras.*, 2014

Automatic Job Ads Classification, Based on Unstructured Text Analysis

Stevan J. Ostrogonac[1], Borko S. Rastović[1], Branislav Popović[2]

[1]Infostud 3 Ltd, Vladimira Nazora 7, 24000, Subotica, Serbia, e-mail: stevan.ostrogonac@infostud.com, borko@infostud.com

[2]Faculty of Technical Sciences, University of Novi Sad, Trg Dositeja Obradovića 6, 21000, Novi Sad, Serbia, e-mail: bpopovic@uns.ac.rs

Abstract: Machine learning models have been tested on countless classification problems in the past. However, there is little information available on how well they perform when the task is learning abstract concepts, that are difficult to understand, even for humans. The object of this research was to find the best model for capturing the concepts of white-collar and blue-collar jobs, based on unstructured job ads, in Serbian. These concepts have become very difficult to define in the modern job market, since there are now many factors besides the required level of formal education, that determines the category of a job.

Keywords: document classification; natural language processing; neural networks; support vector machines; Serbian

1 Introduction

In recent years, different forms and levels of artificial intelligence (AI) have become common tools in all areas of everyday life. In industry, data analysis and machine learning (ML) are used for automation of several important business elements. One of them is personalization of user experience through various recommender systems [1] or general improvement of user experience by using natural language processing (NLP) to make search by keywords more flexible and efficient [2]. Furthermore, data analysis reports can be automatically generated. They provide relevant information based on which project managers can create business strategies and make important decisions. Search engine optimization is another important aspect of business that can benefit from exploiting machine learning algorithms [3]. However, automation of some parts or even entire production processes within enterprises is the most common application of AI.

The research that will be presented within this paper is a part of the efforts aimed at automating the process for the preparation of job ads, that need to be performed prior to publishing the contents, on job boards. More precisely, the research is focused on ads classification to white-collar and blue-collar job positions. Based on this classification, each ad is redirected to a suitable portal within a group of job boards which constitute the website of the company Infostud Ltd.

The problem of textual document classification is one of the main contemporary tasks of natural language processing. Text classification is the key component of spam detection systems [4], automatic user complaints classification for the purpose of forwarding them to relevant persons [5], market analysis [6] and the acquisition of textual data for subsequent usage, which usually includes some machine learning algorithms as well.

In most cases, the document classes are well-defined and intuitive, making it easy for humans to deduce to which class a document belongs to, simply by employing common sense and briefly analyzing textual content. In this research, however, the document classes are not defined by some obvious features. In the past, blue-collar and white-collar jobs were distinguished by the level of formal education needed to perform the duties that those positions imply. In the modern job market, the situation is much more complicated, since there are new positions being created much faster than it is possible to form adequate formal education curriculum. Furthermore, the technology allows people with some basic skills to perform complicated tasks. On the other hand, for some jobs that do not inherently require high levels of education, the opposite applies – the technology dictates the level of skills people need to acquire in order to be able to do their work.

Due to the abovementioned state of the job market, job ads administrators classify the ads based on their empirical knowledge about the difficulty level of skills that are explicitly or implicitly provided in the text. Analyzing the text in order to make this decision takes up significant amounts of time. The administrators also need to consult with one another frequently in order to make sure that the ads will be classified in a unified manner. Business strategies can also influence the classification in some cases. Preparing the data and choosing the right machine learning model, as well as tuning the corresponding hyperparameters for automation of this task or similar tasks, is a special category of problems that have not been extensively explored in the past.

The rest of the paper is organized as follows. In Section 2, the dataset that was used in order to train the models is described. In Section 3, the details on data preparation are provided, including the description of tools and text processing methods which had to be developed and some specifics of the Serbian language which had to be addressed. In Section 4, a simplified description of the system for job ads classification based on ML models is provided. Section 5 contains details related to the experiments. Several types of ML models – Naïve Bayes, Logistic Regression, Multi-Layer Perceptron and Support Vector Machines are examined. Finally, in Section 6, some conclusions concerning the potential of ML for learning abstract concepts are drawn and future research topics are discussed.

2 Data Acquisition and Analysis

Over the past decade, a collection of around 230 thousand job ads have been posted on Infostud job portal poslovi.infostud.com. Roughly 80% of the ads are in Serbian. During most of the above-mentioned period, the ads were manually annotated for market analysis purposes. Among the structured data that were collected, minimal and maximal level of formal education required by employers have been noted, as well as standardized names of job positions, job categories based on fields of work and other. Around 15% of the ads were contained within images, therefore they were not used for training in this research (currently, our text extraction module is not accurate enough).

After selecting the ads suitable for training ML models, the dataset consisted of 130000 ads labelled as white-collar or blue-collar. The labels were inferred from the structured data where possible, and the rest of the ads were annotated manually.

Each ad that was used for the training belongs to one of 55 different areas of work. The distribution of job categories that comprised more than 1% of the entire training corpus is shown in Figure 1. Within the dataset, 56000 different job position names (in further text referred to as positions) were present, each of them being annotated as one of 417 standardized position names. Some of the standardized position names were good indicators for white-collar or blue-collar classification. For example, it was a business decision to treat all information technologies (IT) positions as white-collar, due to the wide variety of technical skills that each of those positions imply. On the other hand, some standardized position names include positions of both blue-collar and white-collar nature, such as pharmacist, which is the standardized position name of both pharmacy technicians and bachelor degree pharmacists.

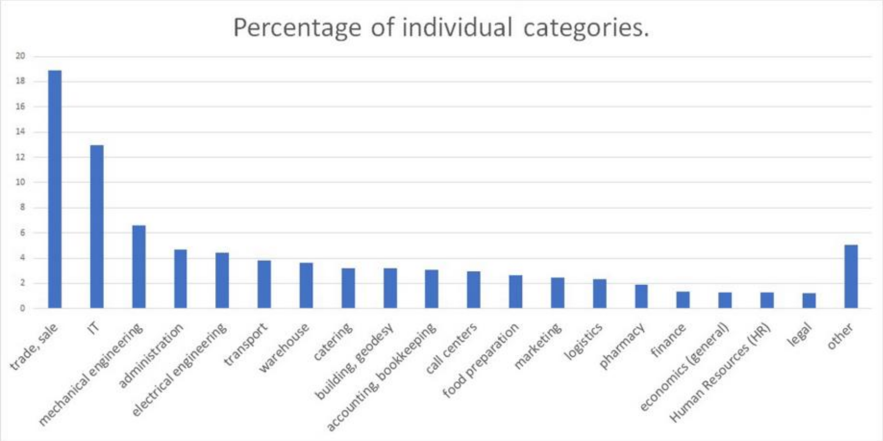


Figure 1
Ad distribution across different fields of work, for the fields that comprise more than 1% of all ads

The most problematic are positions for which it is not clear if they refer to a white-collar or a blue-collar job. Some of these positions are related to standardized position names such as bookkeeper, beautician etc. The ads for these positions would need to be annotated one by one by inspecting textual contents of the ads, which would be extremely time-consuming. Therefore, these ads were not used for the training and the idea was to determine if models can be trained to classify them based on textual data provided from other ads related to similar fields.

Prior to further data preparation, an experiment was conducted in order to determine the importance of word order within these domain-specific texts, which is helpful in choosing the right text representation as input for ML algorithms. Infostud (IS) corpus was compared to a textual corpus that was used for earlier research related to language modeling on the Faculty of Technical Sciences (FTS), University of Novi Sad. Table 1 shows relevant comparative data for the aforementioned corpora. Even though the FTS corpus contains half the number of sentences that comprise the IS corpus, the number of tokens is comparable. This is due to the specific construction of sentences that are typical for job ads that need to reflect short and concise messages. This writing style is similar to journalistic functional style. Extensive studies on the impact of functional styles on the main features of textual corpora have been conducted for Serbian [7-9]. Furthermore, it is an important fact that the vocabulary of FTS corpus is twice as large as the vocabulary of IS corpus. However, the perplexity values are the most indicative. Perplexity on a test set is calculated as the average perplexity on a sentence level, which is calculated as in (1), where *m* is the length of a sentence, and *N* is the order of a language model.

$$ppl = \sqrt[m]{\frac{1}{\prod_{i=1}^{m} P(w_i \,/\, w_{i-N+1},\ldots,w_{i-1})}} \tag{1}$$

Perplexities have been calculated on test portions of each of the corpora (10% of data in both cases) by using previously trained trigram language models. The models were trained using the SRILM language model toolkit [10]. These values show how difficult it is to predict the following word, when previous words are known. The smaller the value, the better the language model. A good language model is an indicator of a high-quality textual corpus. It is important here to emphasize that all the ads form the IS corpus have been reviewed by professional ad administrators, therefore high-quality content is to be expected. From Table 1 it is obvious that IS corpus contains textual content that is much easier to predict. In addition to perplexities, discrimination coefficients have been calculated as well. These values represent perplexities calculated on texts with randomized word orders divided with perplexities calculated on authentic texts [9]. It is evident that the language model trained on the IS corpus can distinguish between natural and random word order very well.

Table 1

Comparative data for textual corpora collected within Infostud and the Faculty of Technical Sciences

	Sentences	**Tokens**	**Vocabulary**	**Perplexity**
FTS corpus	985,498	20,320,616	352,304	294
IS corpus	1,885,625	23,164,550	172,444	16

The above-described experiment indicated that sentence constructions and word orders for job ads are relatively predictable, meaning they do not carry a lot of information. This conclusion implied that the presence of specific words is more important for storing semantic information than their position within textual content. This implied that, within the domain of job ads, bag-of-words (BOW) representation of text [11] might be appropriate for document classification tasks. One alternative to BOW was topic representation [12], which can be obtained by applying Latent Dirichlet Allocation (LDA) [13] to the textual corpus. However, while topic distributions are very useful for classifying job ads based on their fields of work, they do not contain relevant information for classifying the ads to white-collar and blue-collar. Another alternative to BOW was word embedding [14]. Unfortunately, that approach requires a large textual corpus for obtaining vector representations of words, especially for languages with complex morphology, such as Serbian. Therefore, this approach can be taken in the future, after the appropriate training corpus is obtained.

3 Data Preparation

For any machine learning algorithm, it is necessary to prepare data in a way that results in a numerical representation. For the textual content, pre-processing is the first step, performed in order to obtain a sequence of words from the original text, by removing punctuation and other chunks of content useless for a given purpose.

Pre-processing for Serbian is not trivial, since there are no customized tools for performing this task. The most commonly used tools for text pre-processing such as *nltk* library of codes written in Python programming language support many languages, however not Serbian. For this purpose, a Python library called *nlpheart* has been developed within this research [15]. This library, among other features, provides the following text pre-processing steps that are relevant for this research:

- Normalization of symbols (e.g. different forms of quotes need to be replaced with their unique representation)
- Language detection and translation (currently, the language translation module does not produce high-quality results, and therefore it is used only in the prediction phase; in the training phase, only language detection is used to remove content written in languages other than Serbian)

- Separation of words from punctuation (here, various special cases have to be addressed, e.g. web and e-mail addresses, filenames which contain extensions and telephone numbers need to stay in their original forms etc.)
- Conversion of text written in Cyrillic into Latin
- Sentence splitting (not needed for the machine learning algorithm used in this research, but it was needed for the experiment conducted in order to choose text representation model)
- Removal of non-word chunks of text
- Personal data anonymization

After pre-processing, conversion of text into a numerical representation can be performed. For this task, a vocabulary needs to be defined. The vocabulary should include as many of the words that hold useful information for a specific task, but it should not be too large, in order to avoid 'the curse of dimensionality' [16]. The domain-specific vocabulary for Serbian has been extracted from the training corpus. All the words that appeared less than 9 times in the corpus (an empirically determined threshold) have been removed, leaving the resulting vocabulary of 40,758 words. Furthermore, a set of 415 stop-words has been extracted and it was later used in the experiments in order to determine the impact of presence and absence of stop-words on the accuracy of models.

The conversion of each ad (document) to its numerical representation was done by representing the document as a vector which was the size of the vocabulary. Each value within the vector represented the number of times a word with the corresponding index in the vocabulary appeared in the document (BOW model). The vectors corresponding to the documents were, naturally, sparse. Even though there are more advanced numerical representations of text, this is still a common way of data preparation for ML in enterprise systems. The described conversion of text to its numerical representation was performed by using *CountVectorizer* class from a Python library called *scikit-learn*, which is one of the commonly used tools for machine learning, especially for natural language processing [17]. One additional step was performed, in order to improve the quality of data representations (although, not for all the experiments, as will be explained in the following section). This step aimed at giving the words that appear in a small number of ads higher significance and giving lower significance to the words that appear in many different ads, since it is probable that those words are not as important for ad classification purposes. This is performed by employing term frequency – inverse document frequency (tf-idf) algorithm [18], which is implemented within the *scikit-learn* library's *TfidfTransformer* module. The default settings were used as implemented in the library (*l*2 norm etc.)

After all the described steps, the data is finally ready for training the models that will be used for ads classification. It should be noted that the class labels need to be converted to numerical forms as well. In this case, 0 and 1 were enough to map the corresponding blue-collar and white-collar classes.

4 Ads Classification System

In this section, a brief description of the system for ads classification that is based on this research and that is currently being used in Infostud is provided in order to better illustrate the contemporary technologies needed for deploying machine learning models, and to point out some of the practical problems that need to be addressed in the process.

The deployment of the classifier was performed in three steps. The first step presumed creating a web application for serving predictions. For this task, the so-called Flask micro framework written in Python was used [18]. Flask is currently one of the most commonly used tools for creating a web service based on machine learning models due to its simplicity. The second step of the deployment include creating a Docker image in order to containerize the environment needed for the service to function on another machine [19]. The third step was running the image to start a Docker container on local Kubernetes cluster.

The above-mentioned technologies allow efficient delivery of software and automate scaling. Furthermore, Flask application allows the models to be loaded into memory once when the application is started, which is important for the system responsiveness. Loading the models for the ads classification system takes about 850 ms, while the time needed to process the input information and return a prediction is between 10 and 30 ms. The entire process, starting with the ad administration tool sending a request to the containerized application via an exposed port and ending with the tool receiving the response that contains the prediction, takes approximately 150 ms, which is acceptable in the sense that the ad administrators do not notice a delay. The schematic diagram of the system is given in Figure 2. The diagram shows that the classification is performed by combining the decisions of two separate models – one that was trained on the textual contents of the ads, and another that was trained on job position names (ad titles). The models are combined in such a manner that, when they give different outputs, final decisions are made by applying some heuristics. This process will be described in detail in the next section.

Each event of an ad administrator changing the decision of the ad classification system triggers the subsystem for monitoring the classifier’s accuracy. The plan is to use the collected data to retrain the models periodically in order to maintain the accuracy, since new jobs are being created and the rules for blue-collar and white-collar class deduction change over time. With this step the entire project of building this enterprise system based on machine learning is completed.

It is important to note that this system can be used for classifying ads in languages besides Serbian, however another service such as Google Translate would be needed. Training separate classifiers for each of the languages in which significant percentage of ads is present would eliminate the need for a translation service, however it would require adequate training datasets.

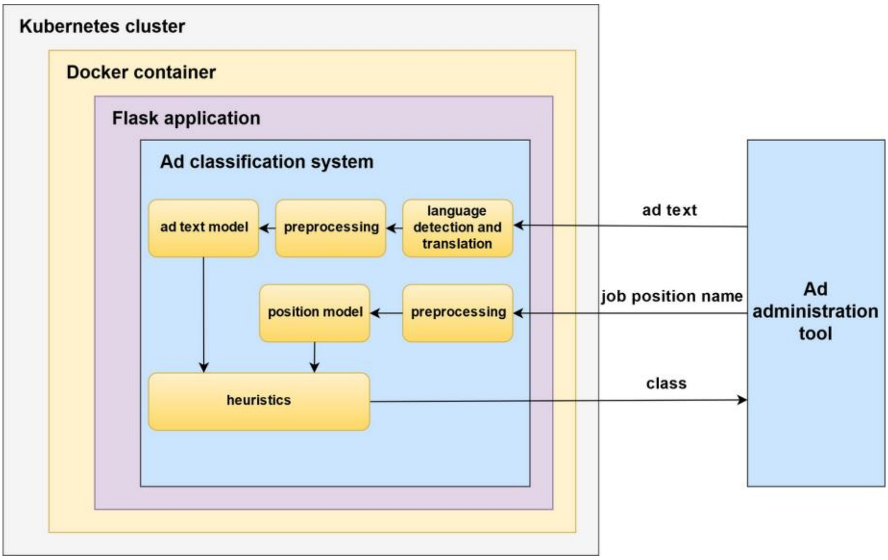


Figure 2
System for job ads classification to blue-collar and white-collar

5 Experiments and Results

The aim of the experiments that will be presented within this section was to test different models and combinations of their respective hyperparameters, as well as, determining how some data preparation steps influence the quality of the resulting classifiers. In other words, in addition to solving the task at hand, the experiments were aimed at obtaining best-practice rules or guidelines for future projects.

The data preparation steps that were tested include the following:

- Usage (or omission) of stop-words in the vocabulary that served for conversion of text to numerical representation
- Usage of the predefined vocabulary vs. the vocabulary inferred from training data
- Usage of higher-order *n*-grams for numerical representation of text (in addition to using the isolated words (unigrams), sequences of words of length 2 (bigrams) and 3 (trigrams) were included in the numerical representation)
- Usage (or omission) of tf-idf scaling of word (*n*-gram) counts

Four types of models were considered in the experiments – Multinomial Naïve Bayes (MNB), Logistic Regression (LR), Multi-layer Perceptron (MLPC) and Support Vector Machine (SVM). The implementations of all the models that have been used are those from *scikit-learn* library [17]. Deep neural networks (DNNs) were initially considered, however the limitations of available hardware and the size of the available dataset indicated that such complex models would not be beneficial.

The test set included the ads that were published on the job portal within two weeks. This set contained 2123 ads, out of which 1450 (68.3%) were blue-collar and 673 (31.7%) were white-collar. The blue-collar and white-collar distribution of the ads form this test set does not reflect the distribution on the entire database, which is close to uniform.

The first set of experiments included training of multiple MNB models. Considering the fact that Naïve Bayes methods are based on the assumption of conditional independence of every pair of features (which would, in this case, mean that the words within a single ad appear in text independently of one another), it could not be expected that these models achieve the highest classification accuracy. However, they were the most convenient for the initial experiments because of their training speed. Therefore, they were used to obtain information related to the effects of data preparation steps on the accuracy of the resulting classifiers. The results of these test are given in Table 2.

It can be concluded that introducing higher order *n*-grams to feature vectors slightly decreases the quality of classifiers, due to a very large number of dimensions of the feature space. Furthermore, applying tf-idf scaling to word counts ads to the model's accuracy significantly, as well as predefining the vocabulary. Finally, excluding stop-words from the vocabulary helps to obtain better models, but this does not apply to the scenarios in which tf-idf is applied and the vocabulary is predefined. The conclusions drawn from this set of experiments have helped to narrow the grid search space for the following experiments. Some of the data preparation steps influence were, however, explored for other models and the conclusions for the experiments with MNB models were confirmed.

The second set of experiments were aimed at finding the best LR model for classification. Different values of the parameter *alpha* (used to compute the learning rate in *scikit-learn* implementation of LR models – *SGDClassifier* with the loss function set to 'log') were tested. The best accuracy was obtained for the value of *alpha* 0.001. Furthermore, the maximal number of training iterations was varied and it was determined that 5 iterations were enough to obtain optimal results. The best LR model reached the accuracy of 78.7%. This is significantly lower than the best accuracy obtained with a MNB model – 82.89%.

Table 2
Accuracy of MNB models trained with different data preparation configurations

n-gram order	tf-idf applied	stop-words excluded	vocabulary predefined	accuracy [%]
1-gram	no	no	no	78.38
			yes	81.62
		yes	no	79.64
			yes	82.23
	yes	no	no	81.57
			yes	82.89
		yes	no	81.62
			yes	82.85
2-gram	no	no	no	73.05
			yes	81.62
		yes	no	74.60
			yes	82.23
	yes	no	no	78.89
			yes	82.89
		yes	no	79.31
			yes	82.84
3-gram	no	no	no	72.30
			yes	81.62
		yes	no	74.60
			yes	82.23
	yes	no	no	78.18
			yes	82.89
		yes	no	78.70
			yes	82.84

The third set of experiments were focused on MLPC models. The neural network structure which was explored consisted of four layers – the input layer (the size of the vocabulary – 40758), two hidden layers and the output layer, which consisted of a single neuron that produces the values 0 or 1, that correspond to blue-collar and white-collar classes, respectively. The number of neurons in the hidden layer varied in order to determine the best structure for this task. Other than that, maximal number of epochs and learning rate have also been tested. The results showed that relatively small variations in accuracy can be accomplished by optimizing the hyperparameters, which can be observed from Table 3. MLPC models generally perform significantly better than the previously examined MNB models.

The last experiment included tests with SVM models. When using the predetermined best-practice configuration related to data preparation, the model's accuracy was 87.91%, which is comparable to the accuracy of the best MLPC model (87.61%). However, the size of the MLPC model was around 160 MB, while the size of the SVM model was around 40 MB, which made the latter better suited for the deployment. Since the test data set contained significantly more blue-collar than white-collar ads, it is important to obtain more detailed information related to the deployed model's accuracy. The weighted precision of the model is 89.86%, and the weighted recall is 87.89%. The most important information can be obtained from the confusion matrix, which looks as in Table 4.

Table 3
Accuracy of MLPC models

maximal number of epochs	size of the first hidden layer	size of the second hidden layer	learning rate	accuracy
10	20	5	0.001	86.05
			0.01	87.28
			0.5	85.67
40	50	10	0.001	86.62
			0.01	87.09
			0.5	87.46
	100		0.001	87.36
			0.01	87.27
			0.5	86.94
	500	20	0.001	84.49
			0.01	87.61
			0.5	86.05

Table 4
Confusion matrix (from the final model, chosen for production)

	Classified as blue-collar	**Classified as white-collar**
Blue-collar ad	1230	220
White-collar ad	37	636

As it was mentioned in previous sections, in addition to the textual content of the ads, position names (ad titles) were also available for the classification task. Around 56 thousand of annotated short phrases were not adequate for training e.g. a neural network, so MNB model was chosen. The data preparation in this case included all the steps that were used in the previous experiments. If the ad text was written in Serbian, the position name was used as is. The best accuracy in the case when ads were classified based only on the position names was 85.3%. Of course, this raised the question about whether we need to model textual content of the ads at all. Therefore, some error analysis needed to be conducted in order to determine the redundancy of these models.

The analysis showed that the model of ad texts misclassified 257 ads, while the model of position names misclassified 264 ads. However, only 145 ads were misclassified by both models. This meant that there was possibility of gaining better accuracy by combining these models and using some heuristics obtained from error analysis to choose which model's decision to consider as final in which situations. After defining a set of those heuristics, the accuracy of the system that combined the two models was 90.9%. The heuristics included the presence or non-existence of certain indicative words or phrases in ads (e.g. managed, director, scientist, professor, pharmaceutical technician, help desk, etc.). The remaining errors were determined to be in fact related to the positions for which there were no annotated data in the training set. These positions mostly belonged to the fields of bookkeeping, sales and work related to beauty or fitness. However, after the inspection of the misclassified ads, it was determined that those were the ads that the annotators were not unanimous about when deciding about the class. The majority of ads in these fields were easy to classify for the annotators upon reading of the entire text, and the model of text handled those ads very well, even though it was only given the examples that belonged to the related, but not similar fields. The final conclusion of this research was that the created system can replace human annotator for the given task.

Conclusion and Further Research

The research that was presented herein was aimed at exploiting machine learning techniques, to automate the process of job ads classification for blue-collar and white-collar ads. The experiments showed that SVM models, as well as, neural networks have the potential to learn from unstructured text and successfully classify textual documents, based on the concepts that humans cannot easily define or explain by creating a set of rules. The research resulted in knowledge about best practices related to document classification for the explored domain. Furthermore, the work resulted in language resources and text processing tools for the Serbian language, for which, there are currently no alternatives.

Further research on this topic will be directed at automating other steps of ad administration, such as, deducing other elements of the structured data, that are later used by search engines, advanced spell-checking tools and discrimination detection algorithms. Further improvement of the models of textual content will include lemmatization of texts and, in later stages, the introduction of semantic classes, for which a large textual corpus in Serbian will have to be collected. After the implementation of all of the most common text processing tools, the authors plan to release *nlpheart* as an NLP library, for the Serbian language, under GNU General Public License.

Acknowledgement

This research was supported by the Innovation Fund of the Republic of Serbia, under the Matching Grants Program for 2020. The authors would also like to express their gratitude towards Infostud 3 Ltd. for providing data and necessary conditions for the presented experiments to be conducted.

References

[1] Kitazawa T.: Sketching Dynamic User-Item Interactions for Online Item Recommendation. In Proceedings of the 2017 Conference on Human Information Interaction and Retrieval. Oslo, Norway. doi: 10.1145/3020165.3022152 (2017)

[2] Pandiarajan S., Yazhmozhi V. M., Praveen kumar P.: Semantic Search Engine Using Natural Language Processing. Advanced Computer and Communication Engineering Technology, Lecture Notes in Electrical Engineering 315, Springer, Cham. doi: 10.1007/978-3-319-07674-4_53 (2015)

[3] Yuniarthe Y.: Application of Artificial Intelligence (AI) in Search Engine Optimization (SEO). Paper presented at the 2017 International Conference on Soft Computing, Intelligent System and Information Technology (ICSIIT), Denpasar, Indonesia, September 26-29, doi: 10.1109/ICSIIT.2017.15 (2017)

[4] Karim A., Sami A., Shanmugam B., Kannoorpatti K., Aalazab M.: A Comprehensive Survey for IntelligentSpam Email Detection. IEEE Access 7, November, doi: 10.1109/ACCESS.2019.2954791 (2019)

[5] Rathore M., Gupta D., Bhandari D.: Complaint Classification Using Word2Vec Model. International Journal of Engineering & Technology, Vol. 7, No. 4, pp. 402-404, doi: 10.14419/ijet.v7i4.5.20192 (2018)

[6] Kuo C., Nagasawa Sh.: Applying Machine Learning to Market Analysis: Knowing Your Luxury Consumer. Journal of Management Analytics. doi: 10.1080/23270012.2019.1692254 (2019)

[7] Ostrogonac S.: Automatic Detection and Correction of Semantic Errors in Texts in Serbian. Primenjena lingvistika (Applied Linguistics), No. 17, pp. 265-278, ISSN 1451-7124, UDK: 81’33 (2016)

[8] Ostrogonac S., Pakoci E., Sečujski M., Mišković D.: Morphology Based vs Unsupervised Word Clustering for Training Language Models for Serbian. Acta Polytechnica Hungarica, Journal of Applied Sciences, Joint Special Issue on TP Model Transformation and Cognitive Infocommunications, Vol. 16, No. 2, pp. 183-197, ISSN: 1785-8860 (2019)

[9] Ostrogonac S., Sečujski M., Mišković D.: Impact of Training Corpus Size on the Quality of Different Types of Language Models for Serbian. Paper presented at the 20th Telecommunications Forum TELFOR-2012, Belgrade, Serbia, November 20-22, ISBN: 978-1-4799-1419-7 (2012)

[10] Stolcke, A.: SRILM – An Extensible Language Modeling Toolkit. In Proceedings of ICSLP 2, pp. 901-904, Denver, USA (2002)

[11] Zhang Y., Rong J., Zhi-Hua Zh.: Understanding Bag-of-Words Model: A Statistical Framework. International Journal of Machine Learning and Cybernetics, 1(1-4), pp. 43-52, doi: 10.1007/s13042-010-0001-0 (2010)

[12] Hingmire S., Chougule S., Palshikar G., Chakraborti S.: Document Classification by Topic Labeling. Paper presented at the International Conference on Information Retrieval (SIGIR 2013), Dublin, Ireland (2013)

[13] Blei D., Andrew Ng., Jordan M.: Latent Dirichlet Allocation. Journal of Machine Learning Research, 3(4-5), pp. 993-1022, doi: 10.1162/jmlr.2003.3.4-5.993 (2003)

[14] Mikolov T.: Statistical Language Models Based on Neural Networks. PhD Thesis. Brno University of Technology (2012)

[15] Stevan O., Borko R., Elizaveta L.: A Python Package for Text Processing for Serbian: nlpheart, Scientific Technical Review, Vol. 70, No. 3, pp. 41-45, doi: 10.5937/str2003041O (2020)

[16] Sarkar D.: Text Analytics with Python: A Practical Real-World Approach to Gaining Actionable Insights from your Data, 2nd edition, ISBN: 9781484243534 (2019)

[17] Pedregosa et al.: Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, Vol. 12, pp. 2825-2830 (2011)

[18] Armash Aslam F., Hawa Nabeel M., Lokhande P.: Efficient Way of Web Development Using Python and Flask. International Journal of Advanced Research in Computer Science, Vol. 6, No. 2. doi: 10.26483/ijarcs.v6i2.2434 (2015)

[19] Babak Bashari R., Harrison J., Ahmadi M.: An Introduction to Docker and Analysis of its Performance. JCSNS International Journal of Computer Science and Network Security, Vol. 17, No. 3 (2017)

IoT and Wireless Sensor Networking-based Effluent Treatment Plant Monitoring System

Md. Saikat Islam Khan[1,2], Anichur Rahman[1,2], Sifatul Islam[1,2], Mostofa Kamal Nasir[1], Shahab S. Band[3,*], Amir Mosavi[4,*]

[1] Department of Computer Science and Engineering, Mawlana Bhashani Science and Technology University, Tangail 1902, Bangladesh, kamal@mbstu.ac.bd

[2] Department of Computer Science and Engineering, National Institute of Textile Engineering and Research (NITER), Constituent Institute of the University of Dhaka, Savar, Dhaka, Bangladesh, {anis_cse, b.khan_cse, and s.islam_cse} @niter.edu.bd

[3] Future Technology Research Center, College of Future, National Yunlin University of Science and Technology, 123 University Road, Section 3, Douliou, Yunlin 64002, Taiwan; shamshirbands@yuntech.edu.tw

[4] John von Neumann Faculty of Informatics, Óbuda University, Bécsi út 96/b, 1034 Budapest, Hungary, amir.mosavi@nik.uni-obuda.hu

Abstract: Contaminated water became a major issue for our country over the last few decades. One of the main reasons behind this scenario is urbanization and industrialization. Every industry should have an Effluent Treatment Plant (ETP) for treating industrial wastewater and safe disposal to the environment. We implement a system that monitors whether an industry uses ETP or not. To monitor ETP, we need to monitor the untreated wastewater quality. The traditional way offers us a method that is time-consuming and inefficient. To solve this problem, we adopt a model based on Wireless Sensor Networking (WSN), which allows us to keep track of the water quality parameters in real-time. This paper proposes a water quality monitoring system that uses WSN and Internet of Things (IoT) based devices to monitor different parameters of water: temperature by a temperature sensor, turbidity by a turbidity sensor, and pH by a pH sensor. Moreover, the microcontroller of Arduino Uno R3 collects the parameter values from these sensors and transmits the values to the IoT based cloud server using the GSM module. The GSM module is also used to alert the supervisors by sending SMS in case of an emergency. Integrating modules such as sensors, Arduino Uno R3, GSM module, enhances the purpose of the desired system. Finally, we calculate the Water Quality Index (WQI) for the pH and turbidity data to report the water quality status. Also, we compare the WQI status with our cloud status, and it shows excellent performance.

Keywords: water quality; wireless sensor networking; IoT; smart sensor; GSM module; Real-time; plant monitoring; artificial intelligence

1 Introduction

Water is one of the most vital assets for humankind. Without water, no plants or animals on earth would survive. The industry has been growing every year on the back of spiraling demand from domestic and export markets. But because of the growing rate of industry in developing countries like Bangladesh, water is constantly being polluted. Water is mostly being polluted because of the industries discharging untreated waste and effluent into the rivers and cannels. Water related diseases cause 3.4 million deaths each year across the globe, according to WHO Water Day Report. Defiled water is also responsible for the degradation of agricultural land, soil fertility loss, and increases pressure on groundwater. About 200 rivers of Bangladesh directly or obliquely received a large amount of untreated industrial wastes. The World Bank claimed that in Bangladesh, approximately $6.5 billion losses due to untreated water, which is 3.4% of the GDP in 2015. There are many factories and industries in our country. According to a Bangladeshi daily newspaper (The daily star), the textile industry will be discharging 203 billion liters of polluted water into the river's water every year from 2021. ETP is one of the best solutions to sanctify untreated water discharged by industries and factories. According to The Daily Star, currently, 5000 ETPs are initiated in factories and industries, which cover approximately 70% of the textile units [1]. It also said that Bangladesh has around 1,200 weaving mills, 5,000 export-oriented dyeing factories, and 450 spinning mills.

We will monitor the quality of water on the industrial water discharged site. To ensure whether the water is contaminated or not, we need real-time data analysis because the sample is continuously changing. If we want to monitor this water through the lab, then the cost will be high, and efficiency will be lower. In modern times, the wireless sensor network is used in many sectors. Wireless sensor networks have received considerable attention not only in environmental sectors but also in industrial sectors [2-6]. WSN provides a massive advantage on cost because the installation and maintenance expenses are low, and the device that we use is cheaper, which required no writing [7-11]. That's why environmental and industrial monitoring largely depends on WSN technology [12-14]. We can apply this technology in water quality monitoring, which will provide us with the best approach to real-time data acquisition, processing, and transmission. In this paper, we proposed a complete WSN water quality monitoring system that will allow us to monitor the ETP. This system consists of a set of sensors such as pH, temperature, turbidity, an Arduino Uno R3 microcontroller, GSM module, IoT based cloud server (Thingspeak). This system measures different parameters of water, such as pH, temperature, and turbidity. In the results and discussion section, we proved that the system has a great prospect in industrial ETP plant monitoring. In this paper, Section 2 discusses the literature survey on surveying water quality. Then, Section 3 illustrates the method we have implemented. After that, our data collection procedure is in Section 4. Section 5 discusses the results obtained through this system and finally, Section 6 brings a conclusion.

2 Literature Survey

In this paper, our fundamental goal is to monitor the effluent treatment plant in real-time using a wireless sensor network. So, our primary concern is to examine water to determine the water quality parameters such as pH, temperature, and turbidity. The following papers proposed various methods to check the quality of water. This study delineated an efficient IoT-based system for measuring water quality by determining temperature, pH, turbidity, and water level [15]. Here, the data is transmitted to a webpage using GPS and GPRS modules. Another work proposed an intelligent sensor interface for industrial WSN in the IoT environment. They monitored water purity in the pond using a light intensity sensor, digital temperature sensor, turbidity sensor by distributing multiple nodes in different areas. They utilized ZigBee wireless communication, a short communication method, for communication purposes [16]. Further, this paper claimed a cost-effective, low-power transmission system for water quality monitoring in lakes around ANNABA reagent. They considered different parameters such as pH, conductivity, temperature, oxygen concentration, measured by other Arduino-based sensors. The authors suggested a personal computer (PC) as a base station and developed a GUI using MatLab software for visualizing data [17]. This study described an automated agricultural monitoring system (iAgriMon) that uses a hybrid IoT and WSN architecture to monitor temperature and humidity parameters in a greenhouse setting [18]. For improving agricultural ecosystems, a web portal is used to analyze the obtained data. In addition, the researchers propose an intelligent irrigation control system that uses wireless sensor networks, a customized server, and a Wemos D1 small (as the main microcontroller) [19]. They applied the T-test statistical tool to see if the results are acceptable. They also compared their method to the traditional approaches, where results suggested that the proposed method outperformed all the other methods. This research described a system that employs WSN technology in the IoT platform for water resource irrigation and proper water resource usage in Precision Agricultural Farming (PAF) [20]. According to IoT-based applications, these studies presented the novel framework based on leading technologies such as Blockchain, Software-defined networks to innovative fields through IoT platforms [21, 22]. Another research claimed that throughput maximization, latency reduction, a high signal-to-noise ratio, a low mean square error, and increased coverage area promote communication between different IoT sensors. They showed that their outcomes outperform traditional IoT-based farming methods [23]. In similar work, Temperature, dissolved oxygen, and pH are measured by the system, which sends the data to a cloud internet platform via a router gateway and can be tracked by smart devices in real-time. This studied offered an upgraded WSN that applies a proposed algorithm in a tomato-growing greenhouse [24]. They monitored Temperature, humidity, Carbon Monoxide, Carbon dioxide, and light intensity and let users set minimum and maximum setpoints and time and date-based irrigation management. They focused on tomato crop yield has grown by 30%, while Methane Gas, water, and electricity use have lowered by 30%, 24%, and 10%, respectively, compared to the conventional method. This research examined an innovative agricultural system based on a WSN and IoT with numerous goals, including implementing a real-time approach to reduce data loss when transmitting and receiving signals and improving the automation system [25]. Furthermore, this research suggested a clustering methodology based on a fuzzy method used in agriculture to increase node density and coverage area and optimize data link and energy consumption. They demonstrated that the strategy is realistic through simulation because it outperforms scalable conditions in the dead final node and half of the node dead [26]. Again, this study deliberated a wireless sensor-based approach named cluster head selection process into the IoT network [27]. They applied the clustering process with low energy consumption to manage the sensor data over the desired IoT network [28]. Furthermore, this paper proposed a system to inspect the quality of water of Lake Victoria Basin (LVB) in real-time by checking water temperature, dissolved oxygen (DO), pH, and electrical conductivity in real-time, using an RF transceiver and GPS receiver for transmitting data in real-time [29]. Similarly, this study proposed a real-time water quality monitoring system for the River Nile by measuring pH, turbidity, and Temperature. They show a prototype that includes a temperature sensor, pH sensor, turbidity sensor, Raspberry Pi, some communication technologies, a dynamic website, and a mobile application for visualization [30]. Additionally, it provided a Water Quality Monitoring System, which uses a turbidity sensor, pH sensor, temperature sensor, and DO sensor integrated into an Arduino Uno board. For transmitting and receiving values, they considered the LoRa module, which uses the LoRa WAN protocol. They visualized their resulting data on the Thing Speak IoT platform [31]. Their work is based on three fundamental blocks: the water quality monitoring stations, the GPRS data transmission modem, and the monitoring station. They employed an A/D converter to convert the analog signal into a digital signal. To read digital data, they use IPC. The data is sent through the GPRS modem to the monitoring station for analysis purposes [32]. Another research [33], a WQM system in which the sensor nodes are placed within the riverbed and data is transmitted to the base station using GPRS. They allow Raspberry Pi over the SPI interface for interfacing ADC. For enhancing efficiency, they use a one-wire communication protocol in temperature sensors. Moreover, a water quality monitoring process uses an uncrewed aerial vehicle (UAV) robot, which used four propellers. This UAV robot is made of an ATMega 2560 microcontroller, a temperature sensor, a ph sensor, a water turbidity sensor, a GPS sensor, a dissolved oxygen sensor, a 3DR radio telemetry transmitter. They developed a GUI at the base station for visualizing the performance of different parameter values of water [34]. Similar work, a system that uses a pH sensing module, Arduino UNO Board, Temperature Sensing Module, and RF Module. They presented Arduino Ethernet Shield as IoT Module, which pushes data to cloud storage. An Android OS platform-based mobile application was developed, which shows the latest Ph value, temperature, and time stamp. A GUI is also extended to show the graphical interpretation of those data for visualizing.

Finally, they compared the proposed system result with a standalone RFID system [35]. In addition, they considered a sum throughput technique to present the joint maximization of energy harvest and information transmission rate where wireless information and power transfer are used to harvest the energy from radio frequency sources [36]. Similarly, they propose an online monitoring water quality system using WSN in Indonesia [37]. They evaluated water quality using Zigbee wireless communication. The pH and turbidity sensors determine whether water quality is good or bad [38]. They observed flood and water quality using IoT. They combined various sensors such as weather monitoring, soil moisture monitoring, and fire alarm to implement their method [39]. They projected an intelligent water quality monitoring system to prevent the contamination of the water. They build a monitoring center where data is analyzed. For data transmission, they used the GPRS method [40]. Furthermore, they planned a low-cost and real-time water quality monitoring system used in remote lakes, rivers, and other water bodies. They proceeded with DO and pH sensors and developed a mobile application to check the system efficiency [41]. On the other hand, these studies focused on machine learning and deep learning methods to classify water quality conditions. They also calculated the water quality index from various sensors [42, 43, 44, 45].

3 Materials and Methods

Wireless sensor networking is used to collect data about various applications, for example, residential security, surveillance, and ocean monitoring. The IoT motivates the rapid advancement of modern wireless telecommunication and is expected to bring avails to a legionary number of application areas, including the industrial WSN systems. The proposed system, "IoT and WSN based Effluent Treatment Plant Monitoring System," performs real-time water quality monitoring. This section represents the structure of WSN and IoT with their corresponding equipment. This system simply includes WSN sensor nodes, Microcontroller, and IoT platforms. Figure 1 expresses the simple way to monitor the water quality, where the wireless sensors module sent the data to the microcontroller module, and the microcontroller module sent the data to the central server.

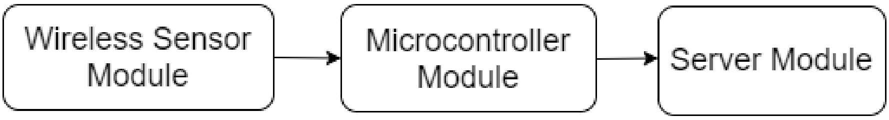


Figure 1

Basic Diagram for Water Quality Monitoring System

3.1 WSN Sensor Node

The sensor node plays the main role in our proposed WSN system. It is sorted with four sensors and microcontroller units. In this system, four sensors measure pH, temperature, turbidity, and flow, which determine the general characteristics of water. However, this method allows using more sensors depending on needs. The pH, temperature, and turbidity sensors are interfaced with the Arduino microcontroller to measure the water quality parameter values. Figure 1 shows the basic diagram of our proposed system, where Figure 2 shows the schematic diagram of the sensor node unit of this system. The sensors are connected to the Arduino Uno board to the correct pin, which ensures correct operation and gives the result of a different parameter value of water correctly.

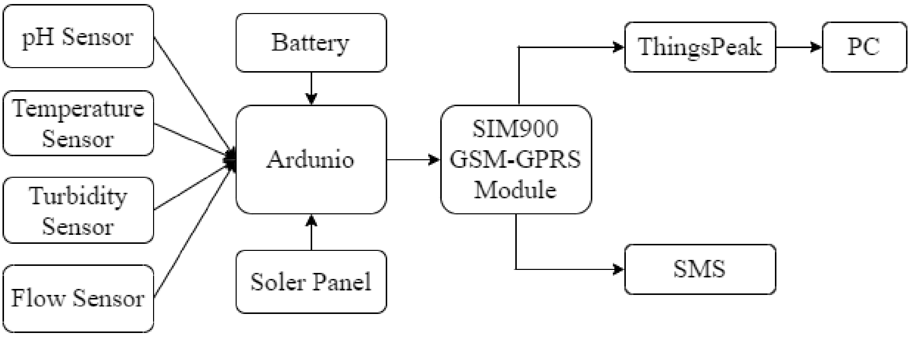


Figure 2
Real-Time Water Quality Monitoring System using WSN

3.1.1 pH Sensor

A pH sensor is a scientific device that is used to measure the hydrogen ion activity in water. The pH sensor determines pH by measuring the voltage level or the difference of the solution in which it is immersed. The logarithmic scale of pH starts from 0 to 14. At level 7, we find the water source level is natural. When the level is less than seven, then the water has acidic solutions, and if the level is greater than seven, then the water has alkaline solutions. A pH sensor got two electrodes, which are the measuring electrode and the reference electrode. The positive end of the battery is paired with a measuring electrode, and a negative end is paired with a reference electrode. The reference electrode will not be changed because it always provides a fixed voltage when the pH meter is dipped into the solutions. The measuring electrode provides voltage and sensitivity to the hydrogen ion. If the temperature changes, then the differential voltage of the electrode also changes.Therefore, we need a temperature sensor.

3.1.2 Turbidity Sensor

The turbidity sensor is the measurement of water transparency. It is used to measure total suspend solids (TSS) in water by sending the light beam into the water body. This light will then fling by any suspended particles such as soil, silts, clay, which enter the water and affect the water body. A light detector is used to measure the amount of light that is being reflected back at it. Turbidity is measured in Nephelometric Turbidity Units, which is known as NTU. Turbidity values from the turbidity sensor can be higher or lower. Higher turbidity means there are lots of suspended solids in the water, and light cannot pass through it, which means the water is impure. Lower turbidity means the water is pure because there are fewer suspended solids in the water, and light can easily pass through it.

3.1.3 Temperature Sensor

Temperature is measured through an electrical signal in the temperature sensor. As the voltage differences of electrodes change with temperature, a temperature sensor is needed. The correction for changing in voltage can also be done by this sensor. It requires RTD (Resistance Temperature Detectors) and a thermocouple. The RTD is a variable resistance that will change the electrical resistance indi-rectly proportional to the change in the temperature in a linear manner where the thermocouple is made by two dissimilar metals which are used to generate the electrical voltage indirectly proportional to the change in the temperature. For pure water temperature value is 27 degrees Celsius. Table 1 represents the pH, temperature, and turbidity value and shows in which condition water is pure or polluted. The general guideline of pH, temperature and turbidity value in pure water is suggested by WHO [46].

Table 1

WSN PARAMETERS VALUE

Parameter	**Treated Water**	**Polluted Water**
pH	6.5-8.5	<6.5 and >8.5
Temperature	20-35 °C	<20 °C and >35 °C
Turbidity	<10 NTU	>0 NTU

3.1.4 Flow Meter

Flow meters measure how much water has gone across it. There are different types of flow meters. Among them, Krohne’s electromagnetic flowmeters can be used to measure both flow volume and flow rate. This flow meter consists of the sensor and converter where the sensor consists of measuring tubes, poles, induction coils, iron core, and shell. It works on the principle of Faraday’s law of electromagnetic induction. Here, digital pulses act as flow volume and can be paired with a microcontroller using a digital I/O pin. We can set 2 flow meters in our proposed system. One in the discharging point of ETP and the other in the discharging point of the water tank. Then, we will compare the total amount of water passed through them. If their volume is significantly different, then an emergency SMS can be sent to the base station.

3.2 Microcontroller Module

Arduino Uno R3 isused as a microcontroller module for this system, which includes a microcontroller and a C program that determines the behavior of the WSN sensor node. Arduino Uno R3 Microcontroller is a free platform that is flexible, convenient hardware, and easy operable software which is used to acquire sensor data. Arduino Uno Board will analyze and process the data and send it to the server. Whenever the pH value goes beyond the 6-8.5 range, Arduino will send an SMS to the authority via the GSM module. Similarly, whenever turbidity value goes beyond the range, as shown in Table 3, the GSM module will send a message about the status of the water quality to the authority.

3.3 SIM900 GSM-GPRS Module

A GSM or GPRS module is a circuit or chip which is used to establish communication between a mobile device or a computer with a GSM or GPRS system. To send the sensors data from the Arduino to the pc, we need a GSM module that is compatible with the Arduino. It allows sending SMS via UART using AT commands. We can create a send SMS() function in the Arduino microcontroller board by using AT commands. This function uses the AT commands such as AT+CMGF=1 ̊and AT +CMGS to send the SMS. A SIM-CARD is attached to the module and is used to send the message to the authority. This module can connect to the internet over the GPRS network. GPRS network provides moderate-speed data transfer using unused time division multiple access (TDMA). SIM900 GSM-GPRS module can transfer sensors data from the Arduino to the IoT cloud server platform using HTTP POST-GET request.

3.4 IoT Platform

An IoT platform is a technology which got more than one layer. It communicates data between a hardware device and cloud storage [47]. Currently, the IoT platform gives users a built-in feature, which makes it easy to create program applications for connected hardware devices, and it also takes care of cross-device compatibility, data security, and scalability. Key technologies that are related to the IoT are sensor node technologies, including wireless sensor networks, miniaturization, and nanotechnology. As IoT is related too many wireless sensor devices, it produces a large number of data, which is also processed by IoT. Basically, IoT consists of three layers 1) Application layer 2) Network layer 3) Perception layer [48]. The data acquisition interface is designed by the perception layer of IoT, which includes sensors, cameras, RFID readers, and various data collection terminals. The data acquisition interface plays a vital role in the collaboration and integration of environments and for collecting the sensor's data. Effluent Treatment Plant can be monitored by the water quality monitoring method. Water quality monitoring is one of the major IoT application fields because it adopts sensors to determine the water quality factor value and detect pollution. That collection of sensors data can be transmitted to the IoT cloud server using the GSM module.

3.4.1 Thingspeak

There are so many IoT platforms that we can use to store, process, and analyze the sensor's data. Some of the IoT platforms are Microsoft Azure IoT,Amazon Web Service or AWS, Google Cloud Platform, Thingspeak, Thingworx, Cisco to IoT CloudConnect, etc. Among them for our method, we use the Thingspeak cloud server, which is an open data platform and API for the IoT, which will help you to store, collect, process, analyze and act on data from sensors. It is also user-friendly and provides data security and free access to the cloud. The sensor node will send data to the cloud to store in the channel of Thingspeak. Thingspeak channel supports eight channels in which we use three channels, such as pH, Temperature, and turbidity. Through this process, we can analyze, visualized, and calculate new data and also interact with social media. The data are coming from the sensors organized in the cloud in the form of plots, charts, graphs using analytical tools online. Thingspeak also provides access to MATLAB to provide sensor data. One can react both in new data and the raw data in each channel and also can help the devices to execute by using the commands. Thingspeak cloud server can send the data to the PC in an EXCEL form which is real-time. The collection of the sensor data in real-time is shown in the data collection section.

3.5 WSN Power Supply

In the WSN system, sensor nodes are situated at a remote distance. So the power supply becomes a major issue here. There are many methods to power the sensor node. Using a battery is one of them. But the battery energy is limited, and replacing batteries is not easy. Different energy harvesting methods like solar panels can be used to recharge the battery. This system uses a 3.7 V 6 AH rechargeable polymer-lithium-ion battery, which is used to power the sensor nodes. This battery has a longer lifespan and also has an excellent self-discharge rate. We will use a solar panel in the future to recharge the battery so that it can save battery power.

3.6 WQI Calculation

A detailed method of WQI calculation can be found [49], but a brief discussion of this method can be found here. To evaluate water quality, we use the scale, which is proposed by Ramakrishnaiah et al. [50] that is shown in Table 2.

Table 2
Water quality scale based on WQI

Water Quality Class	WQI value
Excellent	<50
Good	50-100
Poor	100-200
Very poor	200-300
Unsuitable	>300

WQI can be calculated using the following equation (Brown et al. 1970) [51]

$$WQI = \sum_{j=1}^{n} \frac{wiqi}{\sum wi} \tag{1}$$

Where, wi=unit weight of jth water quality parameters, qi= Quality rating for the jth parameters.

For calculating the WQI, four steps are required.

Step 1: We have selected two variables pH and turbidity for calculating the WQI. We use the standard value of the water quality recommended by WHO [46].

Step 2: Quality rating(qi) can be calculated using the following Equation 2.

$$\mathrm{qi} = \frac{(\mathrm{Va-Vi})}{(\mathrm{Vs-Vi})} \times 100 \tag{2}$$

Where, qi =Quality rating for the jth water quality parameters, Va : The monitored value of the (jth) parameterat a given sampling station, Vi : ideal value for the jth parameters. For pH, the ideal values is 7.0 and for the turbidity variable the ideal value is 0, Vs : the standard value of jth parameters.

Step 3: Unit weight (wi) can be calculated using the following Equation 3.

$$w = \frac{K}{Si} \tag{3}$$

Where, w : unit weight for the jth water quality parameters, Si : standards value for the jth water quality parameters, K: relative constant.

Step 4: The calculated WQI values are classified into five groups. Good water quality is given a low range, and bad water quality is given high range of WQI value.

4 Results Analysis

In this section, we show how we collect sensors data from our experiment. We also show a hardware simulation of how the temperature, pH and turbidity sensors are connected to the Arduino Uno R3 board. For collecting the data from the sensor node, we use a Thingspeak IoT cloud server. The addition of the GSM module allows the system to be more robust and flexible. GSM module allows the sensors to send data to the IoT cloud server. The data collection procedure is shown in Figure 3.

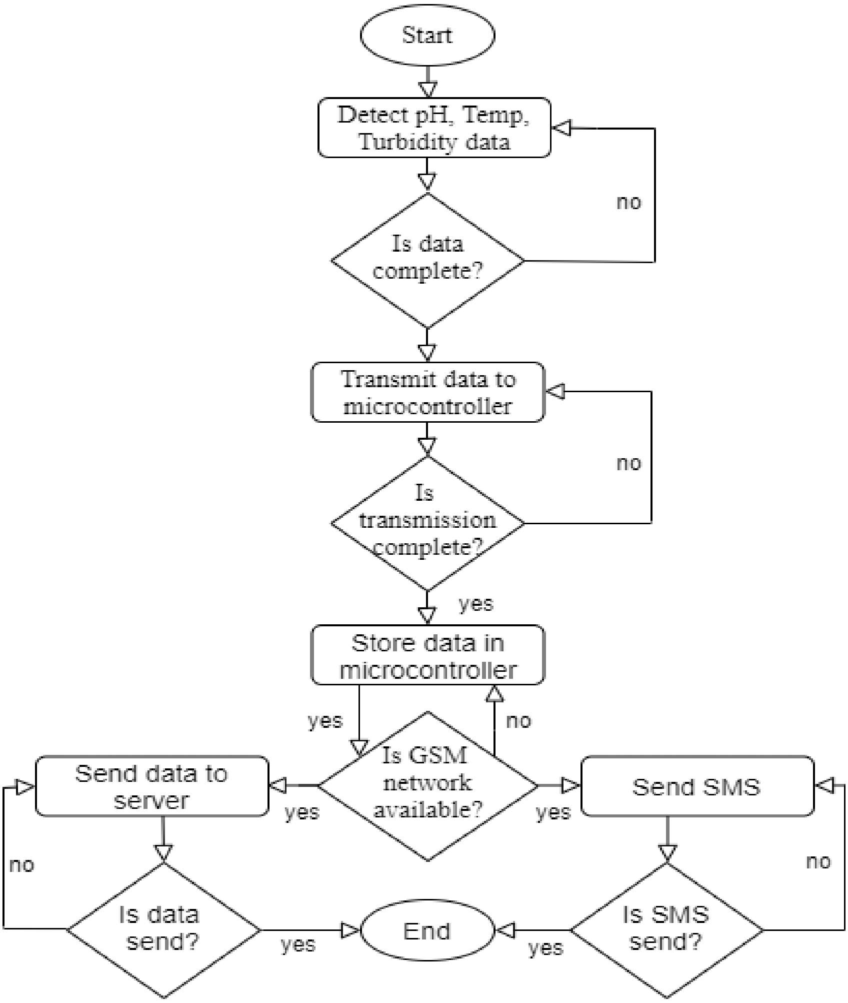

Start
Detect pH, Temp, Turbidity data
Is data complete?
no
Transmit data to microcontroller
Is transmission complete?
no
yes
Store data in microcontroller
yes
no
Is GSM network available?
Send data to server
yes
Send SMS
yes
no
Is data send?
yes
End
yes
Is SMS send?
no

Figure 3
Data Collection Procedure

The optimum pH range for treated and untreated water are shown in Table 1. IoT cloud server will send the data to the corresponding PC in real-time, which is shown in Table 3. Table 3 shows the CSV file generated by the Thingspeak server once data from the GSM module is retrieved. Because of the Thingspeak server-internal mechanisms, the header of this CSV file is fixed to field1, field2, and field3.

Here, field 1 indicates pH (ranges from 0 to 14), field 2 indicates temperature (in degrees Celsius), and field 3 indicates turbidity (in Nephelometric Turbidity Unit).

Table 3
Experiment Value

created_at	entry_i	field 1	field 2	field 3	status
2020-01-24 11:41:00 +06	1	9.242	26.095	0	Treated
2020-01-24 11:46:00 +06	2	9.344	25.73	0	Treated
2020-01-24 11:51:00 +06	3	9.444	26.277	0	Treated
2020-01-24 11:56:00 +06	4	9.648	25.182	0	Treated
2020-01-24 12:01:00 +06	5	9.545	25.73	0	Treated
2020-01-24 12:06:00 +06	6	9.495	25.365	0	Treated
2020-01-24 12:11:00 +06	7	9.091	25.182	0	Treated
2020-01-24 12:16:00 +06	8	8.99	24.818	445	Untreated
2020-01-24 12:21:00 +06	9	8.889	24.635	620	Untreated
2020-01-24 12:26:00 +06	10	8.485	24.635	1942	Untreated
2020-01-24 12:31:00 +06	11	7.929	24.635	3	Treated
2020-01-24 12:36:00 +06	12	8.232	24.453	1093	Untreated
2020-01-24 12:41:00 +06	13	8.737	24.635	2204	Untreated
2020-01-24 12:46:00 +06	14	8.737	39.599	3036	Untreated
2020-01-24 12:51:00 +06	15	8.333	43.613	3004	Untreated
2020-01-24 12:56:00 +06	16	7.727	28.102	0	Treated
2020-01-24 13:01:00 +06	17	9.293	25.365	1518	Untreated
2020-01-24 13:06:00 +06	18	9.404	24.635	0	Treated
2020-01-24 13:11:00 +06	19	9.646	24.635	1685	Untreated

From this Table, we show that data is coming to the PC in real-time. The graphical representation of pH, temperature, and turbidity value is shown in the discussion section. For hardware simulation, we use Fritzing software, which is used primarily for performing schematic capture and allow us to simulate the circuit we design. Using Fritzing software, we simulate how the LM35 temperature sensor, SEN0161 pH meter, and SEN0189 turbidity sensor sends data to the ArduinoUno R3 board. The simulation process is shown in Figure 4. In the simulation process, we con-nect the LM35 VCC pin to the +5V of the Arduino board. As the LM35 output pin produces analog data, so this pin is connected to the 'A1' pin of the Arduino Uno Board. This pin will allow receiving analog values from an exterior origin. The other pinis paired with the GND of the Arduino Uno board. Since pH and temperature sensors are also produced output as analog data, the pH sensor is connected to the 'A0' pin, and the turbidity sensor is connected to the 'A2' pin of the Arduino Uno Board. The other pin is connected to Arduino Uno as the LM35 temperature sensor is connected. We use a serial monitor from the Arduino IDE software in PC, which is used for checking the values of temperature, pH, and turbidity. The serial monitor looks like the LCD monitor, which is also used for showing the parameter values. The calculated WQI values for some of the sample data are shown in Table 4. Comparing the result with our cloud status, it shows an excellent result.

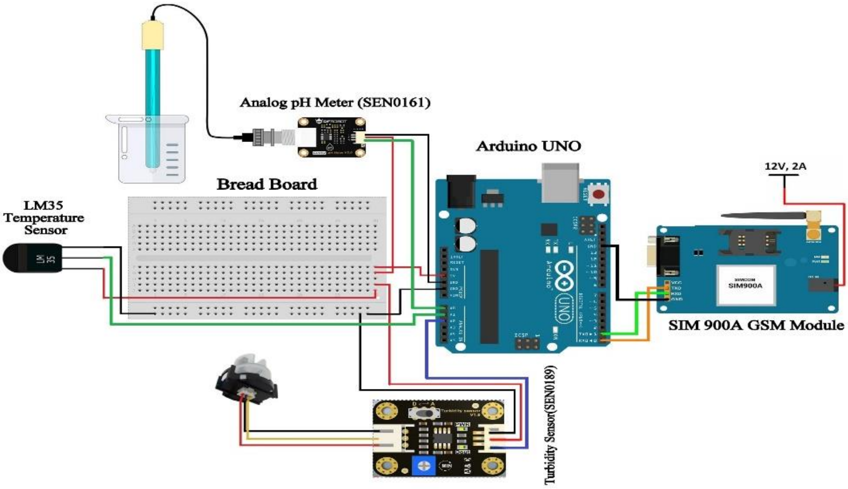


Figure 4

Hardware Simulation Process of Sensors

Although some of the cases can cause problems such as when the cloud server reports that the water quality is treated, but after calculating the WQI, we find that the quality of the water is very poor.

Table 4

WQI values for sample data

pH	Turbidity	WQI	Status
9.242	0	224	Very Poor
7.929	3	106	Poor
8.99	445	2177	Unsuitable
8.485	1942	8780	Unsuitable
7.727	0	73	Good
9.646	1685	7753	Unsuitable

5 Discussion

The proposed method can be implemented in the industry to check water quality in real-time. The results of the method are analyzed and discussed in the context of each scenario. Wireless sensor nodes act as a major role in the whole proposed system. Arduino Uno R3 collects the data from the sensor node. Here, we developed a program and uploaded it to the microcontroller, which allows the system to collect data every 5 minutes. Once the data is calculated, the data is passed to the Thingspeak IoT cloud server using a GSM-GPRS model. IoT cloud server will send the data to the PC in real-time. The performance of the pH, temperature, and turbidity sensor data in the Thingspeak IoT cloud server is shown in Figure 5, 6, 7, which represents how data from the sensor node coming to the Thingspeak IoTcloud server in real-time. Compare to the other methods used for water quality monitoring, our method shows an excellent result and proved to be more effective. Here, our fundamental goal was to monitor the ETP. To monitor the ETP, water quality monitoring is also required. Most of the method used previously is based on just water quality monitoring. In industry to monitor industry wastewater, it is completely a different scenario. Here, we have to consider the environment and also calculate the value in real-time, but most of the methods like paper [15], [16], [31], they use light intensity sensor or turbidity sensor for water quality monitoring. In our method, we use the turbidity sensor, which is more effective in the industry and shows the exact result. Another method like paper [31] they use an ESP32 Wi-Fi module to send the data to the IoT cloud server. For the wifi module, it requires a router to get internet, which may not be possible in the industry. But in our proposed system, we use a GSM module where a SIM-CARD is attached. We can send data to the cloud using the GSM module easily. Thi IoT cloud server is connected to the internet, which uploaded the data into the PC. Using the data, the authority could easily monitor the quality of water. In our program, we also build an emergency situation, when the water quality crosses the danger limit, an emergency SMS will be sent to the corresponding mobile phone. The proposed system gives authorities the ability to check water quality parameters at the industrial water discharge site in real-time. Thus authority can easily monitor whether a particular industry discharged water is polluted or not.

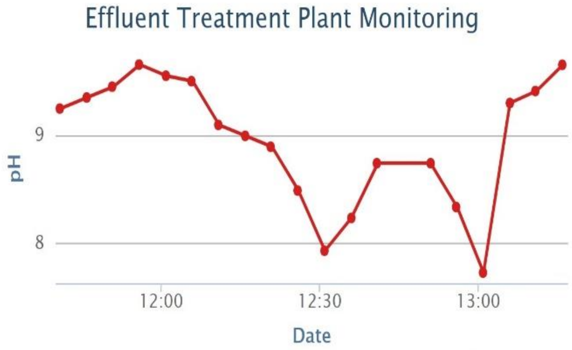


Figure 5
pH Data in Cloud

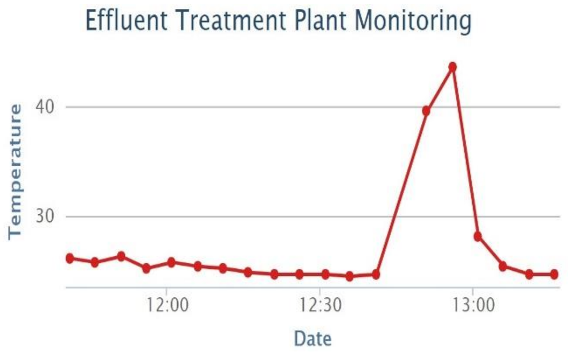


Figure 6
Temperature Data in Cloud

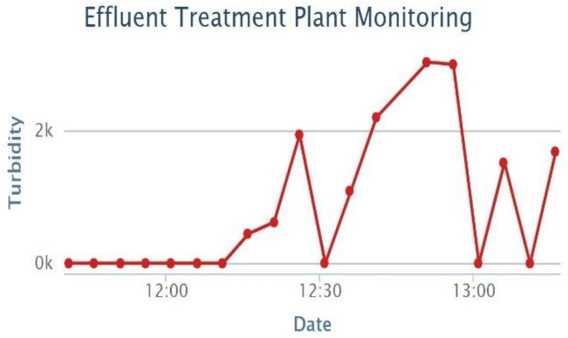
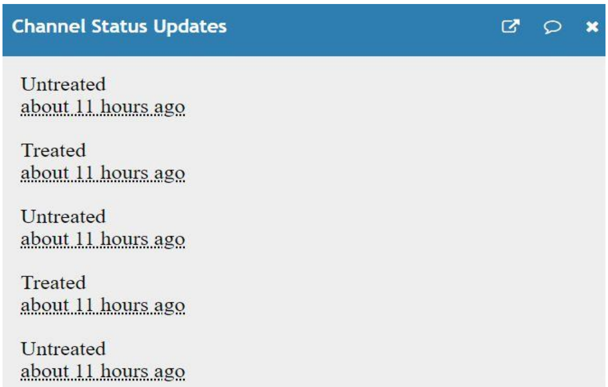


Figure 7
Turbidity Data in Cloud

The status of the data is shown in the Figure 8.



Untreated
about 11 hours ago

Treated
about 11 hours ago

Untreated
about 11 hours ago

Treated
about 11 hours ago

Untreated
about 11 hours ago

Figure 8
Status

Conclusion

In this work, our goal was to monitor the Effluent Treatment Plant using a wireless sensor network in real-time. This proposed system shows an approach for water quality monitoring using a wireless sensor network, which is automated, cost-effective, real-time, server-based, and more effective. We already demonstrated our field test result with appropriate calibration, which proves that the system can monitor water quality parameters tirelessly and send those data to the cloud server. Current procedures that are used in Bangladesh are expensive, non-real-time, and time-consuming. The problems that can affect our system are if the internet speed is slow, then the system will take time to send data. In the future, we will use machine learning to detect whether the water is clean or polluted. We will use more advanced software for our simulation purpose.

References

[1] Rabbi MA, Hossen J, Sarwar M, Roy PK, Shaheed SB, Hasan MM. Investigation of waste water quality parameters discharged from textile manufacturing industries of bangladesh. Current World Environment. 2018;13(2)

[2] Touhami A, Benahmed K, Bounaama F. Monitoring of Greenhouse Based on Internet of Things and Wireless Sensor Network. InInternational conference on the Sciences of Electronics, Technologies of Information and Telecommunications 2018 Dec 18 (pp. 281-289) Springer, Cham

[3] Gomathi N, Jagtap MA. Smart Agriculture System Towards Iot Based Wireless Sensor Network. Turkish Journal of Computer and Mathematics Education (TURCOMAT) 2021 Jun 5;12(6):4133-50

[4] Rajput A, Kumaravelu VB. Scalable and sustainable wireless sensor networks for agricultural application of Internet of things using fuzzy c-means algorithm. Sustainable Computing: Informatics and Systems. 2019 Jun 1;22:62-74

[5] Gidlund M, Han S, Sisinni E, Saifullah A, Jennehag U. Guest editorial from industrial wireless sensor networks to industrial internet of things. IEEE Transactions on Industrial Informatics. 2018;4(5):2194-8

[6] Kiani F, Seyyedabbasi A. Wireless sensor network and internet of things in precision agriculture, Vol. 99, 2018, pp. 99-103

[7] Manrique JA, Rueda-Rueda JS, Portocarrero JM. Contrasting internet of things and wireless sensor network from a conceptual overview. In 2016 IEEE international conference on Internet of Things (iThings) and IEEE green computing and communications (GreenCom) and IEEE cyber, physical and social computing (CPSCom) and IEEE smart data (SmartData) 2016 Dec 15 (pp. 252-257) IEEE

[8] Boonnam N, Pitakphongmetha J, Kajornkasirat S, Horanont T, Somkiadcharoen D, Prapakornpilai J. Optimal plant growth in smart farm hydroponics system using the integration of wireless sensor networks into internet of things. Adv. Sci. Technol. Eng. Syst. J. 2017;2(3):1006-12

[9] Zulkifli CZ, Noor NN. Wireless Sensor Network and Internet of Things (IoT) Solution in Agriculture. Pertanika Journal of Science & Technology. 2017 Jan 1;25(1)

[10] Halim, A. A. A., Hassan, N. M., Zakaria, A., Kamarudin, L. M., Bakar, A. H. A. Internet of things technology for greenhouse monitoring and management system based on wireless sensor network (2016) ARPN Journal of Engineering and Applied Sciences

[11] Zhu J, Song Y, Jiang D, Song H. Multi-armed bandit channel access scheme with cognitive radio technology in wireless sensor networks for the internet of things. IEEE access. 2016 Aug 17;4:4609-17

[12] Xiao-Yan A, Dong-Sheng X, Feng Z, Jian-Gang D. Agriculture intelligent control system algorithm for wireless sensor networks based on internet of things. Sensors & Transducers. 2013 Nov 1;158(11):70

[13] Mainetti L, Patrono L, Vilei A. Evolution of wireless sensor networks towards the internet of things: A survey. In SoftCOM 2011, 19th international conference on software, telecommunications and computer networks 2011 Sep 15 (pp. 1-6) IEEE

[14] Li L, Xiaoguang H, Ke C, Ketai H. The applications of wifi-based wireless sensor network in internet of things and smart grid. In 2011 6th IEEE Conference on Industrial Electronics and Applications 2011 Jun 21 (pp. 789-793) IEEE

[15] M. Parameswari, M. B. Moses, Efficient analysis of water quality measurement reporting system using IOT based system in WSN, Cluster Computing 22 (2019) 12193-12201

[16] Q. Chi, H. Yan, C. Zhang, Z. Pang, L. Da Xu, A reconfigurable smart sensor interface for industrial WSN in IoT environment, IEEE transactions on industrial informatics 10 (2014) 1417-1425

[17] A. Al-Dahoud, M. Fezari, H. Mehamdia, Water Quality Monitoring System Using WSN in Tanga Lake, in: International Conference on Dependability and Complex Systems, Annaba, Algeria, 2019, pp. 1-9

[18] Marques, G., Pitarma, R. An Internet of Things and Wireless Sensor Networks Hybrid Architecture for Precision Agriculture Monitoring (2021) Environmental Science and Engineering, 219, pp. 1863-1867

[19] C. J. H. Pornillos et al., "Smart Irrigation Control System Using Wireless Sensor Network Via Internet-of-Things," 2020 IEEE 12th International Conference on Humanoid, Nanotechnology, Information Technology,

Communication and Control, Environment, and Management (HNICEM), 2020, pp. 1-6, doi: 10.1109/HNICEM51456.2020.9399995

[20] Sanjeevi P, Prasanna S, Siva Kumar B, Gunasekaran G, Alagiri I, Vijay Anand R (2020) Precision agriculture and farming using internet of things based on wireless sensor network. Trans Emerg Telecommun Technol e3978

[21] Rahman, Anichur and Islam, Md. Jahidul and Rahman, Ziaur and Reza, Md. Mahfuz and Anwar, Adnan and Mahmud, M. A. Parvez and Nasir, Mostofa Kamal and Noor, Rafidah Md., "DistB-Condo: Distributed Blockchain-Based IoT-SDN Model for Smart Condominium," in IEEE Access, Vol. 8, pp. 209594-209609, 2020

[22] A. Rahman, M. K. Nasir, Z. Rahman, A. Mosavi, S. S. and B. Minaei-Bidgoli, "DistBlockBuilding: A Distributed Blockchain-Based SDN-IoT Network for Smart Building Management," in IEEE Access, Vol. 8, pp. 140008-140018, 2020

[23] Boonnam, N., Pitakphongmetha, J., Kajornkasirat, S., Horanont, T., Somkiadcharoen, D., Prapakornpilai, J. Optimal Plant Growth in Smart Farm Hydroponics System using the Integration of Wireless Sensor Networks into Internet of Things (2017) Advances in Science, Technology and Engineering Systems, Vol. 2, 2017, pp.1006-1012

[24] Abbasi-Kesbi R, Nikfarjam A, Nemati M. Developed wireless sensor network to supervise the essential parameters in greenhouses for internet of things applications<? show [AQ="" ID=" Q1]"?. IET Circuits, Devices & Systems. 2020 Dec 15;14(8):1258-64

[25] Anulekshmi, S., Durga, R. Comprehensive study and research on wireless sensor network and internet of things for precision agriculture (2020) Journal of Advanced Research in Dynamical and Control Systems, Vol. 12, 2020, pp. 150-158

[26] Yassine, S., Najib, E. K., Fatima, L. Dynamic Cluster Head Selection Method for Wireless Sensor Network for Agricultural Application of Internet of Things based Fuzzy C-means Clustering Algorithm (2019) 7th Mediterranean Congress of Telecommunications 2019, CMT 2019, Faculty of Medicine and PharmacyFez; Morocco; 24 October 2019 through 25 October 2019; Category numberCFP19T86-ART; Code 156056

[27] A. Rahman, M. J. Islam, F. A. Sunny and M. K. Nasir, "DistBlockSDN: A Distributed Secure Blockchain Based SDN-IoT Architecture with NFV Implementation for Smart Cities," 2019 2nd International Conference on Innovation in Engineering and Technology (ICIET), 2019, pp. 1-6

[28] Islam, M. J., Rahman, A., Kabir, S., Khatun, A., Pritom, A., & Chowdhury, M. (2021) SDoT-NFV: A Distributed SDN Based Security System with IoT for Smart City Environments. GUB Journal of Science and Engineering, 7, 27-35

[29] A. Faustine, A. N. Mvuma, H. J. Mongi, M. C. Gabriel, A. J. Tenge, S. B. Kucel, Wireless Sensor Networks for Water Quality Monitoring and Control within Lake Victoria Basin, Scientific Research Publishing Inc. 6 (2014)281-290

[30] N. Kamal, A. Hammad, T. Salem, M. Omar, EARLY WARNING AND WATER QUALITY, LOW-COST IOT BASED MONITORING SYSTEM, Journal of Engineering Sciences Assiut University Faculty of Engineering 47(2019) 796-808

[31] K. Simitha, S. Raj, IoT and WSN Based Water Quality Monitoring System, in: 2019 3rd International conference on Electronics, Communication and AerospaceTechnology (ICECA), Coimbatore, India, 2019, pp. 205-210

[32] T.-z. Qiao, L. Song, The design of multi-parameter online monitoring system of water quality based on GPRS, in: 2010 International Conference on Multimedia Technology, Ningbo, China, 2010, pp. 1-3

[33] S. Doshi, S. Dube, Wireless Sensor Network to Monitor River Water Impurity, in: International Conference on Computer Network and Communication Technologies, Pune, India, 2019, pp. 809-817

[34] B. Etikasari, S. Kautsar, H. Riskiawan, D. Setyohadi et al., Wireless sensor network development in unmanned aerial vehicle (uav) for water quality monitoring system, in: IOP Conference Series: Earth and Environmental Science, Politeknik Negeri Jember, Indonesia, 2020, p. 012061

[35] K. H. Kamaludin, W. Ismail, Water quality monitoring with internet of things (IoT), in: 2017 IEEE Conference on Systems, Process and Control (ICSPC), Melaka, Malaysia, 2017, pp. 18-23

[36] S. O. Olatinwo, T.-H. Joubert, Optimizing the energy and throughput of a water-quality monitoring system,Sensors 18 (2018) 1198

[37] Salim TI, Alam HS, Pratama RP, Anto IA, Munandar A. Portable and online water quality monitoring system using wireless sensor network. In 2017 2nd International Conference on Automation, Cognitive Science, Optics, Micro Electro-Mechanical System, and Information Technology (ICACOMIT) 2017 Oct 23 (pp. 34-40) IEEE

[38] Suryawanshi V, Khandekar M. Design and development of Wireless Sensor Network (WSN) for water quality monitoring using Zigbee. In 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS) 2018 Jun 14 (pp. 862-865) IEEE

[39] Jegadeesan S, Dhamodaran M, Shanmugapriya SS. Wireless Sensor Network based Flood and Water Quality Monitoring System using IoT. Taga journal of graphic technology, Online ISSN. 2018(1748-0345)

[40] L. N. Devi, G. K. Reddy, A. N. Rao, Live demonstrationon smart water quality monitoring system using wireless sensor networks, in: 2018 IEEE SENSORS, New Delhi, India, 2018, pp. 1-4

[41] A. T. Demetillo, M. V. Japitana, E. B. Taboada, Asystem for monitoring water quality in a large aquatic area using wireless sensor network technology, Sustainable Environment Research 29 (12)

[42] U. Ahmed, R. Mumtaz, H. Anwar, A. A. Shah, R. Irfan, J. Garc ́ıa-Nieto, Efficient water quality prediction usingsupervised machine learning, Water 11 (2019) 2210

[43] S. Hafeez, M. S. Wong, H. C. Ho, M. Nazeer, J. Nichol, S. Abbas, D. Tang, K. H. Lee, L. Pun, Comparison of machine learning algorithms for retrieval of water quality indicators in case-ii waters: a case study of hong kong, Remote sensing 11 (2019) 617

[44] Khan MS, Islam N, Uddin J, Islam S, Nasir MK. Water Quality Prediction and Classification Based on Principal Component Regression and Gradient Boosting Classifier Approach. Journal of King Saud University-Computer and Information Sciences. 2021 Jun 14

[45] R. Barzegar, M. T. Aalami, J. Adamowski, Short-term water quality variable prediction using a hybrid cnn-lstm deep learning model, Stochastic Environmental Researchand Risk Assessment (2020) 1-19

[46] Guidelines for drinking-water quality, third edition Edi-tion, World Health Organization, Geneva, 2004

[47] A. Rahman, M. J. Islam, M. Saikat Islam Khan, S. Kabir, A. I. Pritom and M. Razaul Karim, "Block-SDoTCloud: Enhancing Security of Cloud Storage through Blockchain-based SDN in IoT Network," 2020 2nd International Conference on Sustainable Technologies for Industry 4.0 (STI), 2020, pp. 1-6

[48] Rahman, Anichur and Islam, Md. Jahidul and Montieri, Antonio and Nasir, Mostofa Kamal and Reza, Md. Mahfuz and Band, Shahab S. and Pescape, Antonio and Hasan, Mahedi and Sookhak, Mehdi and Mosavi, Amir, "SmartBlock-SDN: An Optimized Blockchain-SDN Framework for Resource Management in IoT," in IEEE Access, Vol. 9, pp. 28361-28376, 2021

[49] M. Kachroud, F. Trolard, M. Kefi, S. Jebari, G. Bourri ́e,Water quality indices: Challenges and application limits in the literature, Water 11 (2019) 361

[50] C. Ramakrishnaiah, C. Sadashivaiah, G. Ranganna, Assessment of water quality index for the ground water in tumkur taluk, karnataka state, india, Journal of Chemistry 6 (2009) 523-530

[51] Brown RM, McClelland NI, Deininger RA, Tozer RG. A water quality index-do we dare. Water and sewage works. 1970 Oct; 117(10)

Analysis of the Relation between State of Health and Self-Discharge of Li-Ion Batteries

Milán Attila Sőrés, Bálint Hartmann

Department of Electric Power Engineering BME Faculty of Electrical Engineering and Informatics, Egry J. u. 18, 1111 Budapest, Hungary
sores.milan@vet.bme.hu, hartmann.balint@vet.bme.hu

Abstract: Li-ion batteries have become a widespread solution for modern energy storage systems, both for e-mobility and stationary storage. SOC and SOH estimation of batteries has great importance from both technical and economic aspects. There are many ways to estimate SOC and SOH with different complexity and accuracy rates, our paper focuses mostly on SOH. At first, this paper gives a brief review of possible methods of SOH determination. From these methods, one way is developed to measure an indicator related to SOH, and from the indicator estimating it. In our paper, we analyzed the connection between SOH and self-discharge for different time periods. The capacity degradation was measured with a high current, that closely resembles modern e-mobility applications. After that, from our experimental data, with the measured self-discharge, the final best-estimated SOH value in the range of ± 3% is achieved.

Keywords: battery; Li-ion; state-of-health; self-discharge; degradation

1 Introduction

Nowadays, it is becoming a more and more important task to estimate battery cell SOC (state-of-charge) and SOH (state-of-health), both in e-mobility and stationary energy storage systems. The main focus of these is application dependent, though there are some basic criteria in accuracy, speed, and robustness. It is application dependent as in the case of SOC there are several electrochemical-based differential equations accurate enough, but they cannot be implemented on microcontrollers or FPGA due to their limited calculation capacity. There are simplified equations based on equivalent circuits or coulomb counting algorithms combining SOC estimation with other measurable quantities such as OCV (open-circuit voltage). However, some basic difficulties can be handled differently. First of all, there is the problem with the non-linearity of the SOC-OCV relationship that can be solved either by different methods, like the Kalman-filter-based method [1] or piecewise linear interpolation method [2]. An issue with these methods is that they may require information on the individual cell that is not available as manufacturers keep them as an industrial secret. From an economic point of view battery degradation cost shall be considered in the design phase of an energy storage system as shown in [3].

Battery health is usually connected with capacity fade and internal impedance increase, in the case of direct measurements. For capacity measurements, a precise current measuring device is required, as the SOC is an open-loop integral. In addition, the measurements shall be carried out with low current in terms of C-rate. Impedance measurements depend on the used battery model. One of the most complex ways is the electrochemical impedance spectroscopy (EIS) that gives the impedance values for a great range of frequencies [4]. In the case of a 2nd order, that was presented in [5] and used in many applications for example in [6]. Considering two RC branches the question is the determination of R_0, R_1, and R_2 (R_1 and R_2 are sometimes named short and long referring to the time constant). In the advanced model, the hysteresis of these values is also considered, which means these values are dependent on the current direction (different resistance for charge and discharge). An indirect often used method is based on the investigation of OCV or pseudo-OCV curves. OCV is the no-load voltage of a battery, after a given relaxation period. This measurement takes time, therefore sometimes a so-called pseudo-OCV curve is used, constant current discharge with current below C/25. All these measurements require time and a proper testing device, for application the latter is sometimes not available, and in this case, the self-discharge measurement could present an acceptable estimation of the SOH.

This paper has the following structure, following the Introduction. In Section 2, State-of-Health Estimation, existing SOH estimation methods will be presented. Then in Section 3, Self-Discharge is presented, the authors present the most common literature concepts for self-discharge modeling and present a model for which the estimation will be applied. Following, in Section 4, Measurements experimental setup will be presented including the measurement devices and the selected cells and the measurement process will also be discussed. After this, in Section 5, the measurement results will be presented and evaluated in Section 6, the Measurement results. From the measurement results, a method for simplified capacity estimation will be discussed in the following Section 7, Estimation. Finally, Section 8, provides the Conclusions of our work, with the summarization of the measurements and all estimations.

2 State-of-Health Estimation

There are several possibilities for SOH estimation. In most cases, SOH decrease is connected to either the loss of available capacity or increasing internal impedance (especially in equivalent circuits). Most information comes from the post-mortem analysis of a battery cell [7]. In their analysis, D. Aurbach et. al examined a 18650-size cell from LG chem, with around 1800 mAh capacity [8]. They cycled the cells at two different temperatures (25°C and 40°C) for about 250 and 300 cycles, respectively. At 25°C the cells reached 80% of their capacity after 230-250 cycles, while at 40°C experiments 140-160 cycles according to their results. They observed both the anode and cathodes with SEM (scanning electron microscope) and compared the Nyquist plots of EIS measurements. Four possible reasons were presented for capacity fading: degradation of active mass on the anode; degradation of the solution; degradation of active mass on the cathode; reactions on the surfaces of both electrodes elevating their impedance, hence increasing the impedance of the whole battery cell. Waldmann et. al [9] examined 18650-size NCA/graphite cells with a wider temperature range. Their experiments were conducted at 0°C, 5°C, 25°C, and 45°C. In their study, they recorded the voltage relaxation for 4 hours from the end of the charge. Their discharge method for the 3250 mAh capacity cells used a relatively small discharge current, 0.5C. They find that at low temperatures (0°C) the main aging factor is Li plating, while at higher temperatures (45°C) it is related to SEI growth and adhesive loss of active material. In [10] the authors give a review on calendar aging mechanisms of different cell chemistries. In their paper, they compared calendar capacity loss and resistance increase of the different cell chemistries at different SOC levels and temperatures. Results show that different ambient temperature and SOC level has different effects on different cell types. As an example, for their paper, an LCO chemistry cell tolerates high temperatures from a calendar capacity fade point of view around 80% SOC level. On the contrary with 80% maximum SOC a bigger battery would be required for the same energy or performance. That would mean a serious disadvantage in applications where mass or volume limitations are present for example an aircraft. Unfortunately, post mortem analysis can give details only after the cell died. For similar reasons, other invasive methods cannot be applied as well. For non-invasive methods, there are two major categories either based on voltage signal or non-voltage signals. Voltage signal-based measurement methods typically use OCV, impedance, electrochemical parameters. Other non-voltage signal-based methods are temperature, ultrasound, and force measurement [11]. In [12] authors used a resistance-based two-tier DC load pattern, and a discharge and relaxation method for OCV, and an equivalent circuit real-time estimation method. The discharge phase lasted for one hour, while the relaxation was chosen to be two hours. In their experiments, Lüders et al. [13] has already examined the relation of Li plating and voltage relaxation curves. As indicated previously the Li plating occurs at low temperatures, therefore their experiments were performed at -2°C. For discharge they used low discharge rates for standard CC-CV (constant current-constant voltage) charge varied from 0.05 to 1 C. Voltage relaxation in their measurements was recorded continuously for 4 hours. The main goal in selected papers was to show the main factors affecting the battery SOH, and the methods that help in its estimation. There is another possibility used in measuring

SOH, it is based on the voltage change during discharge, Differential Voltage Analysis, used in [14] to give a mathematical model for health decrease.

Usually, in the previously mentioned methods, the basic concept is to find a measurable quantity that is closely related to the degradation of the battery. In this paper, we observed the self-discharge of a chosen cell that undergoes continuous capacity fade during the first half of the cell lifetime. Capacity fade is reached on one hand with high C-rate discharge [15] and on the other hand with charging and storing the cell around 4.2 V (~100% SOC) [16]. In our measurement, we chose a cell that can be used in e-mobility applications, as there is a need for higher C-rate discharge in these applications. In their review, Barré et. al [17] compare battery aging estimation methods, based on five main aspects (adaptation, precision, operation without data, real-time application, and prediction). The compared methods: direct measurement; equivalent circuit method; electrochemical model; performance models; analytical model and statistical method. From their comparison, direct measurement is the best at adaptation and precision, while in real-time application and prediction is the worst. From the already developed methods, electrochemical models are the most precise, though their adaptation is rather challenging. Statistical methods are the easiest to adapt, while cannot operate without data. The other methods are between these highlighted ones. During the presented measurements, the capacity was directly measured for comparison with the estimated values. In our experiments, we analyzed the connection between self-discharge and capacity fade. With the presented method a new SOH indicating factor is presented, that can be used both for individual cell SOH estimation and estimation of battery packs or swappable battery modules.

3 Self-Discharge

The self-discharge phenomenon is present for all types of batteries such as lead-acid, Ni-MH, and of course for Li-ion. As our measurements were performed with Li-ion cells, their self-discharge will be discussed in short. The mechanism is related to SOC, electrochemistry, electrolyte, chemical reactions, and temperature [18]. From the internal reactions point of view, it is strongly related to the so-called SEI formation [19].

This SEI layer consumes intercalated Lithium both during charge-discharge reactions and self-discharge [20]. SEI formation is in connection with irreversible capacity loss as well. Beyond of SEI layer, in [21] it was discussed how the electrolyte can contribute to self-discharge in the case of a solution of LiPF6 in linear and cyclic carbonates. Besides the reactions of the electrolyte, there are several processes causing self-discharge as well. Reactions such as internal electron leakage, dissolution of active electrode materials, corrosion of current collectors, and parasitic electrochemical reactions on the electrode surface according to [22]. For simulation of self-discharge, there are semi-empirical models developed based usually on the Butler-Volmer equation and Nernst equation, as presented in the overview on the self-discharge topic of the authors [23]. From those methods, the one developed by Galushkin et al. [24] and adapted to Li-ion batteries by Deutschen et. al. [25] is used for estimation. They derived two equations for approximate voltage decay one for short term (1), and one for long term (3), with switching from short term to long term around the 25th day.

$$V_t(t) = V_{t,0} - \frac{1}{b} \cdot \ln\left(1 + \frac{I_0 \cdot b \cdot \exp(b \cdot V_0) - 1}{C} t\right) \tag{1}$$

This short equation is formally very similar to the so-called Nernst equation [26]. It has five parameters, from which *b* can be further expressed (2).

$$b = \frac{zF}{RT} \tag{2}$$

For Li-ion cells $z = 1$ (as the number of electrons involved in the electrode reaction [26]), F is the Faraday constant, R is the gas constant, and T is the absolute temperature. From (2) it is obvious that V_t depends on temperature as well, and with knowledge of the temperature value b can be calculated.

$$V_t(t) = V_{t,\infty} + \frac{1 - \exp(-b \cdot V_0)}{b} exp\left(-\frac{I_0 \cdot b}{C} \cdot t\right) \tag{3}$$

Long-term approximation (3) has similar parameters, the difference is that in (1) an initial voltage is given, while here a stationary voltage is used.

4 Measurements

In this section, the measurement setup, chosen cell, self-discharge phenomenon, and the measurement process are presented. Since the aim of the measurements was to observe the degradation of the chosen battery cells that can be used in e-mobility applications. For this purpose, the cells were discharged with a high constant current rate (4C) as e-mobility applications like electrical aviation requires quite high discharge rates, opposing stationary storage systems, where the load can be rather low [27]. The selected current rate will be further explained in Subsection 2.1 selected cells.

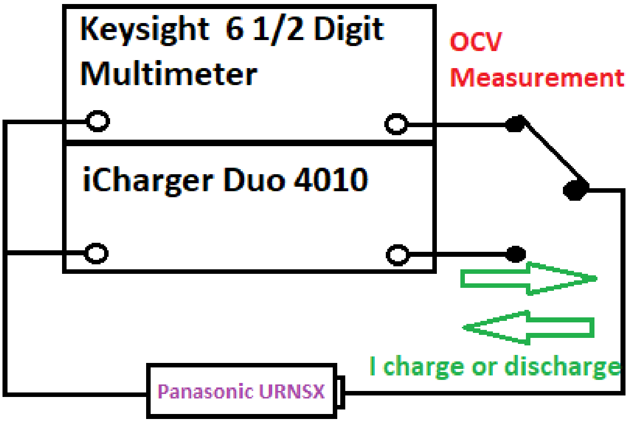


Figure 1
Schematic diagram of the test setup

Figure 1 shows the measurement setup schematic. The tested cells are connected to the iCharger Duo 4010 device for charging and discharging. As the iCharger is capable of voltage measurement as well the digital multimeter is not used during this phase. For open-circuit voltage (OCV) only the 6½ digit Keysight multimeter was used, as it provides more precise values. During the measurements, ambient conditions are recorded as well (e.g. temperature). The list of the equipment can be found in Table 1.

Table 1
Devices used in the measurement

Device	Measurement type
Keysight 34461A (6 ½ digit multimeter)	OCV
Junsi iCharger Duo 4010	Charge and Discharge voltage and current
Testo 622 (Scientific Ambient Monitor)	Ambient conditions

4.1 Selected Cells

The selected cells for the measurements were Panasonic UR18650NSX type [28]. These are cylindrical 18650 size cells with a 2.5 Ah rated capacity. According to its datasheet, a cell can be discharged with continuous 10C, that is 25 A, tough around 20% of SOC the battery temperature reaches 80°C. That is why a 10C discharge rate may not be preferable for e-mobility applications, while in the range of a bit smaller discharge rate (8C) the cell temperature remains below 80 °C. In our test method, 4C constant current discharge was selected, as it is still a high current value (10 A) and the cell delivers performance as well. Beyond power and energy considerations it is also an important factor that the temperature of the cells remains certainly at an acceptable level. The manufacturer provides information on the cyclic performance of the battery cell with 4C and 10C discharge rates. In the first case, the cells reach 80% of their initial capacity during 300 cycles, while in the second case during 200 cycles. This also means that a higher discharge rate has a negative effect on the cycle life of a cell. The reason for choosing this cell is the possibility of high discharge rates that is essential for e-mobility applications where the required performance can be very high. Another aspect of the selection was the relatively high energy density of the cell, approximately 204 Wh/kg gravimetric and 514 Wh/l in volumetric means according to the datasheet.

4.2 Measurement Process

These measurements aimed to examine the effects of this high current discharge on the self-discharge of the cells. Therefore, the open-circuit voltage (OCV) was measured 10 min, 24 h, 48 h, and one week after fully charging the cells. The 10-minute measurements were done after each charging, but for obvious time limitations, the other three (24 h, 48 h, and one week) were done only after 10 cycles. After charging the cell the voltage decreases relatively fast, due to polarization, a 10-minute relaxation time was chosen for OCV measurements similar to [29], where the authors used from 6 minutes to 5 hours relaxation times. In [30] the authors find that the steady-state is reached in the range of 24 h. In the presented measurements one cycle means one full discharge and recharge.

Altogether 5 Panasonic URNSX cells were tested. Originally there were 6 cells but after 35 cycles of testing, one of them had to be terminated due to some mechanical damage. The mechanical problem was independent of the test procedure. Altogether 100 full cycles were done.

The test contained three major parts:

- Charging the cells to 100% SOC (CC-CV)
- Relaxation
- Discharging with 4C until 2.65V (CC)

During the initial phase, before the first discharge, of the test, cells were charged to maximum SOC, and for a week. During this initial relaxation period, the OCV of the cell was measured after 10 min, 24 h, 48 h, and a week. Then another charging started to compensate for the slight capacity loss due to self-discharge. The next step was the discharge with 4C, after that charging again and relaxation for 10 min, and OCV measurement. These were repeated 10 times, that is 10 cycles altogether. After the tenth cycle, a longer relaxation period started for a week, measuring voltage the same way in the initial phase. The charging step was performed according to CC-CV protocol, wherein CC phase 1C (2.5 A) was used. In the CV phase the target voltage was 4.2 V. In cases of charge after the relaxing period, only the CV phase was performed. For discharge, as mentioned previously, a 4C constant current was used with cut-off voltage 2.65 V as a safety limit as opposed to the 2.5 V. During the discharge phase cell voltage [V], current [A], and relative time [s] were recoded, while capacity [mAh], power [W], and energy [Wh] were calculated from them. During each test step, ambient conditions were recorded. These are: temperature [°C], relative humidity [%] and air pressure [hPa].

5 Measurement Results

5.1 Capacity Measurement

During the capacity measurements, the original six cells were cycled. They were labeled from S10 to S15. Measurements of S14 were terminated, due to some mechanical damage, not related to the charge-discharge cycles. In this section, some graphs from the discharge measurements are shown. Capacity was calculated by the Coulomb counting (or Amper counting) method. The method is described by (4) based on [31]:

$$SOC = SOC_0 - \frac{1}{Q_r} \cdot \int_{t_0}^{t} \eta I(\tau) d\tau \tag{4}$$

According to (4) the state of charge consists of two parts SOC_0, the initial SOC (in the presented case always 100% at the start of discharge. The other term is an integral of the charge or discharges current between the initial time t_0 and t. As SOC is a percentage the result of the integration is divided by the term Q_r, the rated capacity from the datasheet. Since the capacity of a new cell usually differs from Q_r, because of different ambient conditions for example; and in our measurement, the used discharge was not the datasheet specified method (not measured at 20 °C) therefore the initial capacity was used through the whole measurement. The constant η is called the coulombic efficiency, for Li-ion batteries, it is very close to 1 [32], therefore it is considered to be constant 1. In this paper the capacity of the cell is denoted by Q, to avoid any confusion with the current rate C.

Fig. 2 shows the voltage of one battery cell during 4C constant current discharge during the first 20 cycles. The voltage maximum during charge was set to 4.2 V, the initial voltage at the start, after 10 minutes relax, was between 4.189 V and 4.178 V depending mostly on the ambient temperature. The cut-off voltage was set to 2.65 V during all discharge cases to ensure a safer operation and not to reach deep discharge accidentally. In practice under about 3 V with load, the cell voltage drops more and more fast. At this range, the remaining capacity is relatively small, while the temperature increases fast. This temperature rise can be explained by the increasing serial resistance at low SOC values from the equivalent circuit point of view. From Fig. 2 it can be seen that at the first 20 cycles the cell performance slightly decreased, as at the first cycles the discharge took about 830 seconds, while at the end of the sequence it took only 820 seconds. The nominal time for 4C discharge would be 900 secs or 15 minutes. The difference from the ideal quarter-hour measurement time, is probably because the voltage did not reach 2.5 V and the cells themselves have some difference from catalog values.

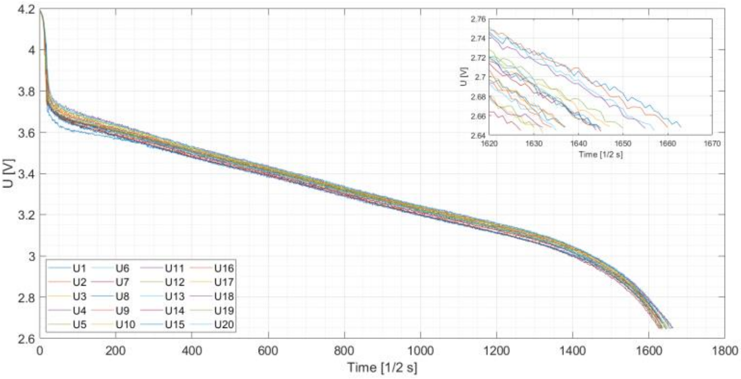


Figure 2
Voltage vs. time of one cell during the first 20 cycles

On the aging of the cells, not the first 20 cycles would give information but all the cycles. Since the one-week relaxation occurred only after 10 cycles, therefore, the measurements after these relaxation periods are shown in Fig. 3.

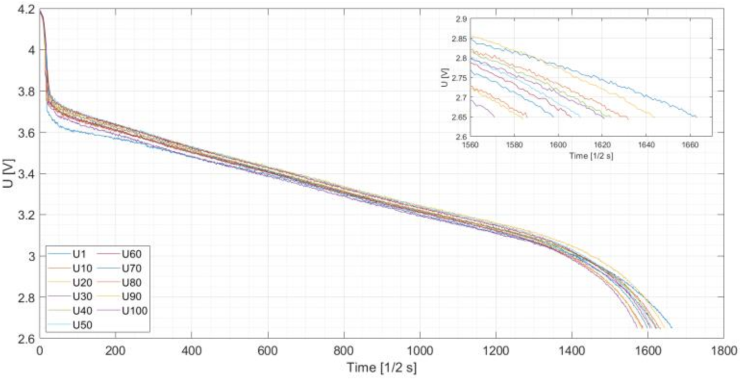


Figure 3
Cell Voltage of one cell during cycling, only every 10th cycle presented

Voltages at end of the measurement, only the last 150 seconds, are shown in the upper right corner of Fig. 3 as well. It shows that at the first cycle, with fresh cells the discharge took the most time, from 10th to 40th the time is very similar to each other. From the 50th, the second half of our tests it monotonously shortens from around 810 seconds to 785 seconds.

Figure 4 shows the measured capacity in percentage of all the five cells and their average value (red line). The capacity should monotously decrease with cycling, but in Fig. 4 there are some local minimums, after which capacity seems to be growing a bit. These are because of the so-called overhang effect discussed in depth in [33] and [34]. Briefly, as the anode is designed to be slightly wider than the cathode, some Li atoms can diffuse to this region in case of longer storage periods. The apparent capacity loss is slowly reversible with cycling, that is the reason of this curve. The overhang effect is significant in higher SOC ranges.

For simplification reasons, the measured capacity of the first discharge measurement is considered to be 100%. From Fig. 4 and Fig. 5 it can be seen that during the first 15-20 cycles capacity fade has a higher slope after that the slope decreases and becomes almost constant until the 100th cycle. As the five cells follow a very similar pattern in Fig. 5 only the average relative SOC decrease will be presented with a linear approximation of their slope. The yellow line is the approximation of the slope of the first 15 cycles the orange line is the remaining 85 cycles. Their equation is presented on the diagram; their color is respective to the line color. From the equations it can be stated that the slope of the first 15 capacity results is almost 3 times bigger, in absolute value (0.159 [% / cycle]), than the measurements after that (0.57 [% / cycles]).

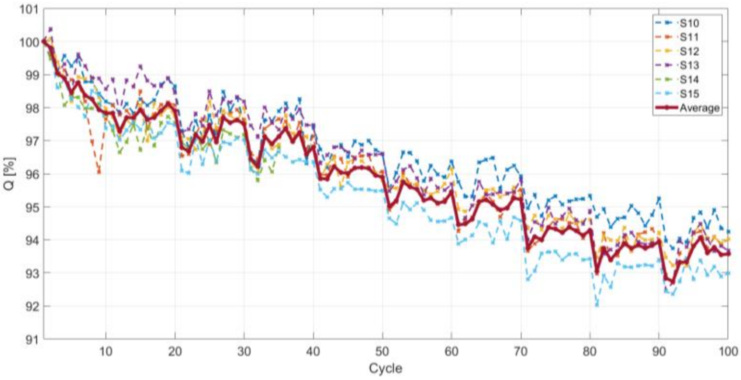


Figure 4
Relative capacity of the cells

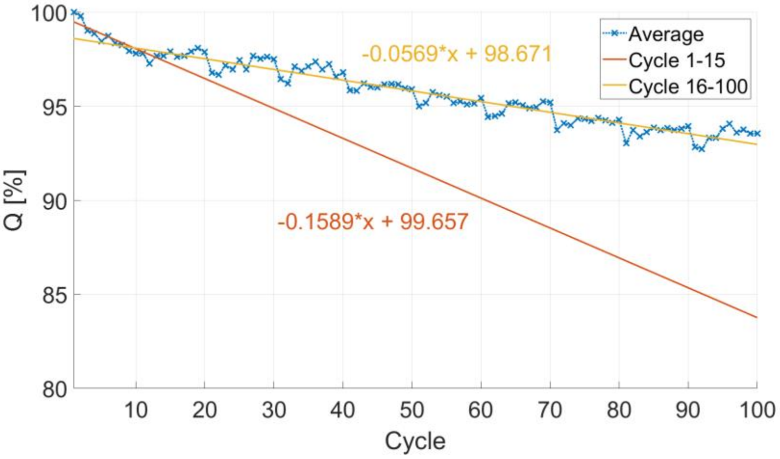


Figure 5
Slope of capacity decrease (first 15 cycles yellow; last 85 orange)

If we neglect the rapid capacity decrease during the few cycles in the beginning it is possible to fit a linear curve to the measured values shown in Figs. 5 and 6. The orange curve was created with the MATLAB curve fitting tool, and the average measured values. The fitting algorithm was set to Linear-Least Squares. The equation of the curve is in Eq. 2.

$$Q(cyc) = q_d \cdot cyc + q_0 \tag{5}$$

In Eq. 2 Q is the capacity, *cyc* is the cycle number, q_d (capacity decrease) and q_0 (initial capacity) are coefficients with values -0.0569, 98.67 respectively, with 95% confidence bounds.

5.2 OCV Measurement

In this section, the results of the OCV measurements will be presented. As described in section 2.3. the 24 h, 48 h, and 1-week measurements were done only after 10 cycles, therefore only 10 measurements results are shown in Fig. 6. For simplification reasons, only the average of the cell voltages will be presented. In addition to the measured values, a linear trendline is presented as well. The trendline was created with MS Excel's built-in function. The measured voltages at 30th and 70th cycles were rather outlier values, probably because of some errors. Therefore, they are omitted in the following estimation methods.

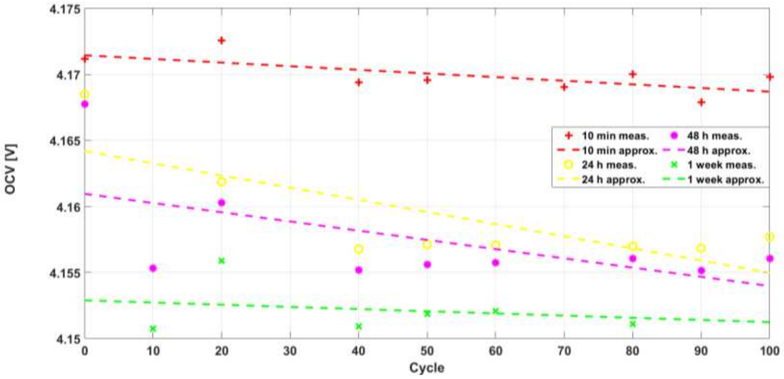


Figure 6

Measured OCV after different relaxation times

It can be seen from the figures that the measured voltages have a decreasing tendency, as expected. The red plus signs and dashed trendline belong to the 10-minute measurements, the yellow circles, and the dashed trendline to the 24-hour relaxation times. Magenta stars and green crosses and dashed linear trendlines belong to the 48 hours and one-week relaxation, respectively. If we compare the diagrams it can be stated that the slope decreases with the relaxation time for 24 h, 48 h, and 1-week OCV data, for 10 min measurements there is a negative slope as well, but it is more similar to the 1-week measurements than to the 24 h or 48 h. The voltage decrease values are summarized in Table 2.

Table 2

Voltage decrease after specified relaxation time

Relaxation Period	v_d [µV/cycle]	v_0 [V]
10 minutes	-27.54	4.171
24 hours	-92.11	4.164
48 hours	-69.84	4.161
1 week	-16.47	4.153

6 Estimation

In the previous section, it is presented that both capacity and open-circuit voltages tend to decrease with cycling. For the voltages, the decrease is different for different relaxation times. This section aims to observe the estimation results for a capacity fade with the help of different voltages. As shown in Fig. 6 the decrease of the average measured capacity values is around 0.057% (relative to the first capacity measurement) per cycle. This information is not available generally, only after long battery cycling processes. On the other hand, manufacturers usually provide some information on cycling characteristics for the battery cell, that can be used for estimation. This is shown in Fig. 7.

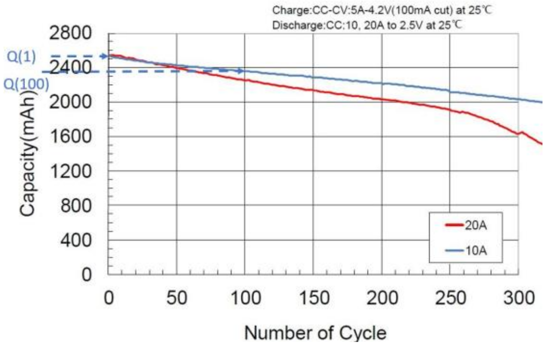


Figure 7

Cycling characteristics of the Panasonic UR18650NSX cell [28]

The manufacturer in this case gave two cycling curves for two different discharge currents. Blue is for 10 A discharge current, which corresponds to 4C, and red curve for 20 A that corresponds to 8C. As our measurements are concluded with 4C, the blue curve is used. From the datasheet curves Q(1), as the initial capacity of the cell and Q(100), capacity after 100 cycles are used. As another point for linear estimation Q(100) is chosen as our measurements lasted only for 100 cycles. From Fig. 7 Q(1) is 2510 mAh, Q(100) is 2360 mAh. Capacity decrease is around 150 mAh for 100 cycles, the q_d value become 0.06 % / cyc. This value aligns with the measured discharge rate quite well.

Voltage decrease will be estimated based on Eq. 1. As previously mentioned the equation has two independent variables *t* and *T*. This would lead to a surface fitting problem. Unfortunately, the measured temperature values are from a very narrow range. The minimum temperature was 23.5 °C, the maximum was 26.6 °C. In terms of *b*, the minimum would be 38.7 the max would be 39.1. That would give questionable results regarding surface fitting. Therefore, instead of using individual *T* values for the fitting, the average will be used, it is 25.5 °C. With this approximation Eq. 1. will be simplified in a way that it contains only two parameters shown in Eq. 5.

$$V_t(t) = V_{t,0} - \frac{1}{b} \cdot (1 + pt) \tag{5}$$

Note that $V_{t,0}$ is not changed, *b* is constant, and *p* incorporates the original parameters I_0, V_0 and, *C*. Simplification was necessary because the original equation continued four parameters (considering *b* constant). The measured voltages after 10 minutes, 24 hours, 48 hours and, 1-week would make it possible to fit the original curve, however, the 10-minute voltage measurement has to be excluded from the curve fitting, as it is not part of the self-discharge period, but the relaxation after current flow. From an equivalent circuit method point of view, only the terminal voltage (V_t) can be measured without disassembling the cell.

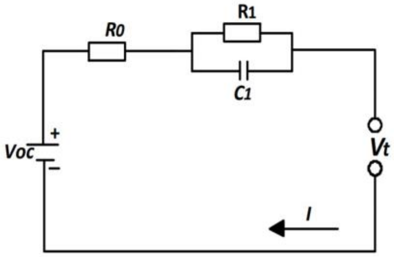


Figure 8
Li-ion battery model with one RC branch [35]

That means measured *Vt* has two terms, open-circuit voltage, and the voltage of the *R1-C1* parallel term. This voltage decreases to zero in time, according to their time constant. A similar idea applies in the case of more parallel RC branches. The time these voltages are negligible depends on the model, can last from minutes to several hours [36].

For estimation and validation the measured cells are split into two sets. Parameter estimation was performed on the first set: S10, S11, S12 cells. These cells are referred as estimation set. The validation was performed on the second set: S13, S14 and S15. Unfortunately, S14 was withdrawn from the measurements, as previously mentioned. These cells are referred as validation set.

With MATLAB Curve Fitting Toolbox, parameters $V_{t,0}$ and p were fit to the measured voltage points. Table 2 shows the parameters, R-square and, RMSE values.

Table 2
Curve fitting results

	$V_{t,0}$	**p**	**R^2**	**RMSE**
10	4.158	0.001722	1	-
20	4.16	0.001722	0.9893	0.000430
40	4.159	0.001825	0.9889	0.000461
50	4.159	0.001582	0.9885	0.000414
60	4.16	0.001517	0.9983	0.000151
80	4.1595	0.001377	0.9391	0.000687
90	4.159	0.001467	0.9659	0.000675

The estimation result shows, that *Vt,0* remains constant, while parameter *p* shows a clear decrease. The 1st measurement is omitted because previous results in Fig. 5 shows that the capacity fade becomes steady between cycle number 10 and 20. For the 100th cycle, the voltage result measurement file was damaged therefore, these values are omitted as well. A similar p estimation was performed for the validation set as well.

Fig. 9 shows the estimated values for p, and a linear trendline is fitted to these values. The slope of the trendline is -4.54 · 10^{-6} (p_d) and the offset is 1.829 · 10^{-3}. These *p* values decline similar to the decline of measured capacity, so the slope of the trendline for p_d and slope of the capacity q_d from the estimation set will be used to estimate capacity fade Eq. 6.

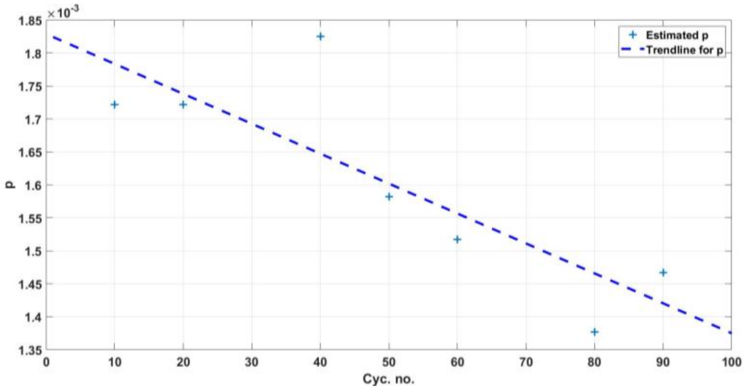


Figure 9

Results of estimating parameter p and trendline fitted to it

Index *v* in Eq. 6 refers to values from the validation set of cells and the init index is the 10th measured value, or fitted in case of *p*.

$$Q_v(cyc) = Q_{v,init} - \frac{q_d}{p_d} \cdot [p_{v,init} - p_v(cyc)] \tag{6}$$

Note that there were only ten voltage measurements for each period, so for p_v(cyc), linear interpolation was applied on the measured data.

7 Estimation Results

Fig. 10 shows the original capacity vs. cycles curve and the estimated capacity curves.

In Fig. 10 the blue points are the measured capacity values of the validation set, the red ones are the estimated capacity values based on Eq. 6.

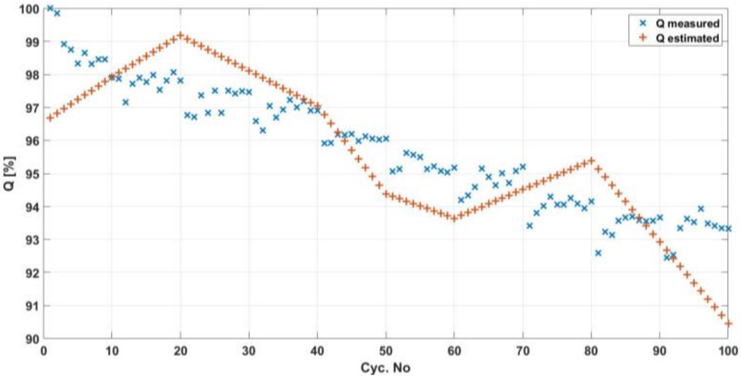


Figure 10

Measured and estimated capacity

Estimation Accuracy

The accuracy of the measurement is presented in Fig. 11. Here the percentage error is calculated by Eq. 7 for the nth cycle.

$$Error(n) = \frac{Q_{measured}(n) - Q_{estimated}(n)}{Q_{measured}(n)} \cdot 100 \qquad (7)$$

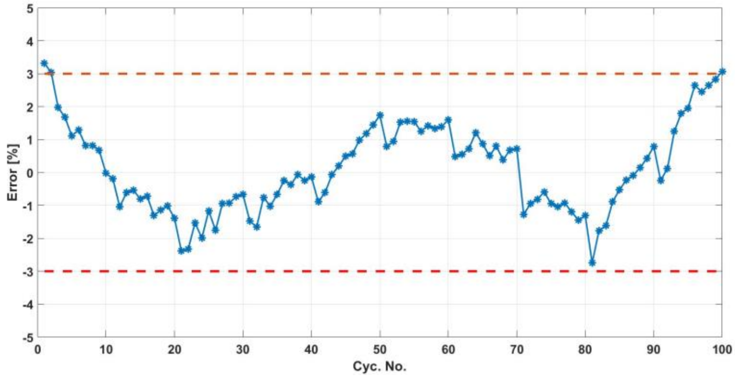


Figure 11

Relative error of the estimated capacities

In. Fig. 11 it is shown that the relative error of the estimation remains in the range of ±3 % for the relevant cycles. Although previously mentioned that the first 10 measurement results were excluded from the estimation process, their estimation result is presented as well in this figure.

Conclusions

In this paper, the authors presented a link between cell SOH (capacity decrease) and the self-discharge process of a chosen Li-ion battery cell, suitable for e-mobility applications. The measurement results involved the first hundred cycles of five cells with a relatively high discharge rate. The main purpose of these tests was to analyze the effects of this high current discharge on the self-discharge of the cells.

Although effects of self-discharge are considered to be smaller for Li-ion technology, the voltage decrease was measurable. Open circuit cell voltage was measured four times after charging was finished: 10 minutes, 24 hours, 48 hours, and one week. A previously developed self-discharge model was applied to investigate aging-related issues. Although a relatively simple estimation method was used for calculating SOH, results show that it can give fairly accurate results. This can be used on battery modules, that are allowed to have a resting period, for example in electric vehicles. All capacity and open circuit voltage measurement data is accessible [37].

References

[1] L. Lavigne, J. Sabatier, J. Mbala Francisco, F. Guillemard, A. Noury: Lithium-ion Open Circuit Voltage (OCV) curve modelling and its ageing adjustment. Journal of Power Sources 324 (2016) 694-703

[2] I. Baccouche, S. Jemmali, A. Mlayah, B. Manai, N. E. B. Amara: Implementation of an Improved Coulomb-Counting Algorithm Based on a Piecewise SOC-OCV Relationship for SOC Estimation of Li-Ion Battery. International Journal of Renewable Energy Research. Vol. 8, No. 1, pp. 178-187, 2018

[3] O. Boqtob, H. El Moussaoui, H. El Markhi and T. Lamhamdi: Energy Scheduling of Isolated Microgrid with Battery Degradation Cost using Hybrid Particle Swarm Optimization with Sine Cosine Acceleration Coefficients. International Journal of Renewable Energy Research. Vol. 10, No. 2, pp. 704-715, 2020

[4] T. Stockley, K. Thanapalan, M. Bowkett, J. Williams, M. Hathway: Advanced EIS Techniques for Performance Evaluation of Li-ion Cells. Proceedings of the 19th World Congress The International Federation of Automatic Control Cape Town, South Africa. August 24-29, 2014

[5] Chen, M., Rincón-Mora, G. A: "Accurate Electrical Battery Model Capable of Predicting Runtime and I–V performance", IEEE Transactions on energy conversion, Vol. 21, No. 2, 2006, pp. 504-511

[6] T. Debreceni, G. Gy. Balázs, I. Varjasi: Mission Profile-Oriented Design of Battery Systems for Electric Vehicles in MATLAB/Simulink. International Conference on Renewable Energies and Power Quality (ICREPQ'16)

[7] A. Friesen, X. Mönnighoff, M. Börner, J. Haetge, F. M. Schappacher, M. Winter: Influence of temperature on the aging behavior of 18650-type lithium ion cells: A comprehensive approach combining electrochemical characterization and post-mortem-analysis. Journal of Power Sources 342 (2017) pp. 88-97, http://dx.doi.org/10.1016/j.jpowsour.2016.12.040

[8] D. Aurbach, B. Markovsky, A. Rodkin, M. Cojocaru, E. Levi a, H.-J. Kim: An analysis of rechargeable lithium-ion batteries after prolonged cycling. Electrochimica Acta 00 (2002) pp. 1-13

[9] T. Waldmann, J. B. Quinn, K. Richter, M. Kasper, A. Tost, A. Klein, M. Wohlfahrt-Mehrens: Electrochemical, Post-Mortem, and ARC Analysis of Li-Ion CellSafety in Second-Life Applications. Journal of The Electrochemical Society, 164 (13) pp. 3154-3162 (2017)

[10] M. Dubarry, N. Qin, P. Brooker: Calendar aging of commercial Li-ion cells of different chemistries – A review. Current Opinion in Electrochemistry Vol. 9, pp. 106-113, 2018

[11] J. Tian, R. Xiong, W. Shen: A review on state of health estimation for lithium ion batteries in photovoltaic systems. eTransportation 2 (2019) 100028, https://doi.org/10.1016/j.etran.2019.100028

[12] Venkatesh Prasad K. S., B. P. Divakar: Real Time Estimation of SoC and SoH of Batteries. International Journal of Renewable Energy Research. Vol. 8, No. 1, pp. 44-55, 2018

[13] C. von Lüders, V. Zinth, S. V. Erharda, P. J. Osswalda, M. Hofmann, R. Gilles, A. Jossen: Lithium plating in lithium-ion batteries investigated by voltagerelaxation and in situ neutron diffraction. Journal of Power Sources, Vol. 342, pp. 17-23, 2017, http://dx.doi.org/10.1016/j.jpowsour.2016.12.032

[14] J. Wang, J. Purewal, P. Liu, J. Hicks-Garner, S. Soukazian, E. Sherman, A. Sorenson, L. Vu, H. Tataria, M. W. Verbrugge: Degradation of lithium ion batteries employing graphite negatives andnickelecobaltemanganese oxideþspinel manganese oxide positives:Part 1, aging mechanisms and life estimation. Journal of Power Sources Vol. 269, pp. 937-948, 2014

[15] G. Ning, B. Haran, B. N. Popov: Capacity fade study of lithium-ion batteries cycled at high discharge rates. Journal of Power Sources 117 (2003) pp. 160-169, doi:10.1016/S0378-7753(03)00029-6

[16] E. Wikner, T. Thiringer: Extending Battery Lifetime by Avoiding High SOC. Applied Sciences 8:1825 (2018) doi:10.3390/app8101825

[17] A. Barréa, B. Deguilhem, S. Grolleau, M. Gérard, F. Suarda, D. Riu: A review on lithium-ion battery ageing mechanisms and estimations for automotive applications. Journal of Power Sources Vol. 241, pp. 680-689, 2013, https://doi.org/10.1016/j.jpowsour.2013.05.040

[18] S. H. Choi, J. Kim, Y. S. Yoon: Self-discharge analysis of LiCoO2 for lithium batteries. Journal of Power Sources 138 (2004) pp. 283-287

[19] C. Wang, X. Zhang, A. J. Appleby, X. Chen, F. E. Little: Self-discharge of secondary lithium-ion graphite anodes. Journal of Power Sources 112 (2002) pp. 98-104

[20] Steven E. Sloop, John B. Kerr, Kim Kinoshita: The role of Li-ion battery electrolyte reactivity in performance decline and self-discharge. Journal of Power Sources 119-121 (2003) pp. 330-337

[21] Ramaraja P. Ramasamy, Jong-Won Lee, Branko N. Popov: Simulation of capacity loss in carbon electrode for lithium-ion cells during storage. Journal of Power Sources 166 (2007) pp. 266-272

[22] Z. Mao, M. Farkhondeh, M. Pritzker, M. Fowler, Z. Chen: Calendar Aging and Gas Generation in Commercial Graphite/NMC-LMO Lithium-Ion Pouch Cell. Journal of The Electrochemical Society, 164 (14) A3469-A3483 (2017). DOI: 10.1149/2.0241714jes

[23] M. A. Sörés and B. Hartmann, Overview of possible methods of determining self-discharge, 2020 IEEE International Conference on Environment and Electrical Engineering and 2020 IEEE Industrial and Commercial Power Systems Europe (EEEIC / I&CPS Europe), Madrid, Spain, 2020, pp. 1-5, doi: 10.1109/EEEIC/ICPSEurope49358.2020.9160787

[24] N. E. Galushkin, N. N. Yazvinskaya, D. N. Galushkin, Generalized model for self-discharge processes in alkaline batteries, J. Electrochem. Soc. 159 (8) (2012) A1315-A1317

[25] Thomas Deutschen, Sophie Gasser, Manuel Schaller, Jochen Siehr: Modeling the self-discharge by voltage decay of a NMC/graphite lithium-ion cell. Journal of Energy Storage 19 (2018) 113-119

[26] J. Yang, C. Du, T. Wang , Y. Gao, X. Cheng, P. Zuo, Y. Ma, J. Wang, G. Yin, J. Xie, B. Lei: Rapid Prediction of the Open-Circuit-Voltage of Lithium Ion Batteries Based on an Effective Voltage Relaxation Model. Energies 2018, 11, 3444; doi:10.3390/en11123444

[27] A. Kumar, A. Biswas: Techno-Economic Optimization of a Stand-alone PV/PHS/Battery Systems for very low load Situation. International Journal of Renewable Energy Research. Vol. 7, No. 2, pp. 844-856, 2017

[28] https://datasheetspdf.com/pdf-file/1160668/Panasonic/UR18650NSX/1 (accessed 2020.12.07)

[29] M. Petzl and M. A. Danzer, Advancements in OCV Measurement and Analysis for Lithium-Ion Batteries, in IEEE Transactions on Energy Conversion, Vol. 28, No. 3, pp. 675-681, Sept. 2013, doi: 10.1109/TEC.2013.2259490

[30] S. Abu-Sharkh, D. Doerffel: Rapid test and non-linear model characterisation of solid-state lithium-ion batteries. Journal of Power Sources Vol. 130, pp. 266-274, 2004

[31] Y. Zheng, M. Ouyang, X. Hanb, L. Lub, J. Lib: Investigating the error sources of the online state of charge estimation methods for lithium-ion batteries in electric vehicles. Journal of Power Sources Vol. 377, pp. 161-188. 2018

[32] H. Wenzl, Batteries and fuel cells, Efficiency, Encyclopedia of Electrochemical. Power Sources, Elsevier, Amsterdam, 2009, pp. 544-551

[33] Balazs Gyenes, D. A. Stevens,V. L. Chevrier, J. R. Dahn: Understanding Anomalous Behavior in Coulombic Efficiency Measurements on Li-Ion Batteries

[34] M. Lewerenz, J. Münnix, J. Schmalstieg, S. Kabitz, M. Knips, Dirk Uwe Sauer: Systematic aging of commercial LiFePO4|Graphite cylindrical cells including a theory explaining rise of capacity during aging. Journal of Power Sources. Vol. 345, pp. 254-263, 2017

[35] B. V. Rajanna, M. K. Kumar: Comparison of one and two time constant models for lithium ion battery. International Journal of Electrical and Computer Engineering Vol, 10, No. 1, pp. 670-680, 2020

[36] M. Messing, T. Shoa and S. Habibi: Lithium-Ion Battery Relaxation Effects. 2019 IEEE Transportation Electrification Conference and Expo (ITEC), pp. 1-6, doi: 10.1109/ITEC.2019.8790449, 2019

[37] https://bitbucket.org/milansores93/dataset_sd_soh/src/main/Measurement_data.xlsx (accsessed 2021.08.10)

Simplified Computation of The Heat Transfer Coefficient in Quenching

Imre Felde

John von Neumann Faculty of Informatics, Óbuda University, Budapest, Hungary,
felde@uni-obuda.hu

Abstract: Due to the direct observation of the Heat Transfer Coefficient at the surface of the metallic components under quenching is practically impossible, indirect methods based on measuring the cooling curves in certain points inside the workpieces, and numerical integration of their thermodynamic model mean a viable approach. The complexity of the necessary calculations can be considerably reduced by the use of symmetric cylindrical samples made of an alloy of particularly simple thermal properties defined in the standard ISO 9950. In spite of that the complexity is still large enough. In the present approach it is reduced by applying a simple formal, qualitative model of the cooling process, the efficient Newton-Raphson algorithm and a Fixed Point Iteration approach to obtain approximate preliminary results. This approach requires only very limited computational capacities. Besides for making rough estimations, due to its simplicity, the method may be useful in the education.

Keywords: Heat Transfer Coefficient; Newton-Raphson Algorithm; Banach's Fixed Point Theorem; Fixed Point Iteration; Quenching

1 Introduction

In many technical applications rigorous deduction of the results on the basis of the available physical models and fundamental principles is impossible due to the high complexity of these tasks. In quenching of metallic components this complexity often roots in the complicated initial and boundary conditions, the turbulent flow of the cooling fluid, phase transitions in the fluid as condensation, evaporation that brings about heat insulating gas bubbles at the surface of the quenched components. Phase transitions and simultaneous chemical reactions within the solid sample may serve as heat sources or sinks, etc. To partly reduce these complexities, in the standard ISO 9950 [1] a cylindrical sample of well defined shape, size, and material composition is applied with internal points in which the temperature can be measured by thermocouples. The task can be formulated as the determination of the *"Heat Transfer Coefficient" (HTC)* on the surface of the sample as the function of the time, in the possession of the measured cooling curves, the heat conduction model, and the main thermodynamic data of the component that are available for some key engineering alloys [2]. If this task is tackled by assuming explicit *HTC*-time functions over the surface of the sample, finite element mathematical approximation and numerical integration of the equations becomes possible to determine the cooling curves. The so computed curves can be compared with the measured ones, and by modification of the original assumption on the *HTC*, efforts can be done to reduce these differences (e.g. [3]).

Even if the cylindrical symmetry of the probe defined in [1] is taken into account, the solution of this *"Inverse Heat Transfer Problem"* needs great computational power. For instance in [4] gradient-based and genetic algorithm was applied with complementary utilization of the graphical card's computational capacities of the computers.

As an alternative of using high computational power in technical applications the systematic utilization of certain *quantitative numerical values* and some available *qualitative modeling knowledge* can be mentioned that at first obtained rigorous mathematical framework in the theory of fuzzy sets [5]. By the combination of the qualitative and quantitative information simple functions with appropriate *"shape parameters"* can be defined, and these parameters can be fitted to the measured data for obtaining good numerical approximations. In Dombi's *"Pliant Systems"* introduced in [6] the so-called *"distending function"* is used due to which the various operators are closely related to each other by setting certain form parameters. Similarly, *"computationally cheap approximation"* of important probability distribution functions (the *normal*, *epsilon*, *omega*), and the *kappa regression function* by typical functions of certain form parameters can be found (e.g. [7]). Similar situation can be observed in modeling dynamical phenomena having long memory properties: they can be described only by very high order differential equations containing numerous parameters. However, by using *fractional order models* (the brief history of the topic is given in [8]) instead of the integer order ones, only a few free parameters can be well fitted to the practical problems.

In the present paper a similar approach is outlined by observing the *qualitative formal properties* of the cooling curves and the calculated HTC vs. time functions given in [9].

For the approximation of these structures it is assumed that at the surface of the sample the function HTC vs. time can be expressed as $HTC \equiv h(T(\mathbf{r},t))$ (variable $t\,[s]$ means the time, vector $\mathbf{r}$ denotes the location of the point on the surface of the probe, and $T\,[C]$ is the temperature). For the single variable $h(T)$ functions various formal suggestions are given. Assuming that the length of the probe is long enough for making the problem symmetric to its centreplane at the half length, the heat transfer in the longitudinal direction must be zero. Furthermore, by utilizing the cylindrical symmetry of the probe, the simplified distribution function $T(r,t)$ was considered, in which $r\,[m]$ denotes the radial distance from the centreline. The appropriate "format functions", the finite element model and the Euler-integration applied, as well as the methods used for tuning the parameters are discussed in Section 2.

2 Formulation of the Problem for "FEM" Approach

For obtaining curves similar to that given in 2 simple qualitative physical considerations were done. Quenching of metal products is a complex process that is difficult to precisely describe by physical models. The *HTC* at the surface of the component is a parameter that may depend on the nature of the bulk flow (laminar or turbulent), the temperature, density, viscosity, the latent heat of evaporation/condensation, and other thermal properties of the cooling liquid, the surface quality, shape, and thermal data of the component under quenching. It is mainly determined by physical processes as follows: i) Even if the "bulk" of the cooling liquid has turbulent flow that allows very efficient heat transfer due to "stirring" the fluid layers, at the boundary layer of the quenched sample laminar flow is formed since the fluid sticks on the surface of the probe. The thickness of this layer is determined by the quality of the surface of the probe, the density and viscosity of the liquid. In this layer the heat transfer process mainly is realized by heat conduction, so it is not very efficient. ii) The viscosity of a liquid normally decreases with increasing temperature, therefore at higher temperatures thinner films with better heat transfer abilities are expected. iii) When the temperature achieves the boiling point of the liquid at the pressure of the operation, at the surface of the probe gas bubbles appear that act as "heat insulators", so at higher temperatures some decrease in the *HTC* value is expected. This effect clearly can be observed optically in the case of refrigerators where it is a practical experimental possibility to manufacture the pipes of transparent glass (e.g. [10]). Though the temperature ranges are considerably different, the main physical processes essentially are the same. It can be commonly observed that boiling liquid drops can have relatively long persistence on hot metallic surfaces. iv) When the hot probe is immersed into the cooling liquid, this liquid comes into boiling nearby its surface. Though the latent heat of boiling used to be considerable, the heat insulation made by the gas bubbles normally has more significant effect on the *HTC*, consequently, in the beginning low *HTC* values can be expected.

The above argumentation well explains the shape of the curves given in [9]. On this reason simple form functions that potentially are able to model the "asymmetries" in $h(T)$ were investigated as given in (1) in which the parameters h_{max}, w, T_{max}, w_{left}, and w_{right} individually must be set to produce values of order of magnitude represented in [9]. In (1a) the formal asymmetry was taken from Planck's radiation law using the frequency as independent variable. (Certain modification was introduced to eliminate the numerically inconvenient behaviour of the original formula at $T = 0$ that may disturb the numerical calculations.) In the other formulae the location of the *"centre of asymmetry"* is determined by the parameter T_{max}, and its extent depends on the *"width parameters"* w_{left} for the lower, and w_{right} for the higher temperatures. The parameters given in Table 1 were set by the method of *"generate and test"*.

Due to the cylindrical symmetry of the problem the use of polar coordinates was expedient that yields the heat conduction equation for the alloy Inconel 600 as given in (2) in which $t\,[s]$ denotes the time, $r\,[m]$ denotes the radius from the centerline as the "independent variables" of the problem, $T\,[C]$ is the temperature, and it is "formally" taken into consideration, that according to the numerical data published on the thermal properties of this metal in [2], within the temperature range $T \in$

$[27,796.45]$ $[°C]$ no observable “source” or “sink” terms appear.

$$h(T) = \frac{h_{max}T^3}{\exp(T/w)-1} \text{ modified as } h(T) = \frac{h_{max}T^3}{\exp(D+|T/w|)-1} \tag{1a}$$

$$h(T) = h_{max}\begin{cases} \exp\left(-([T-T_{max}]/w_{left})^2\right) & \text{if } T \le T_{max} \\ \exp\left(-([T-T_{max}]/w_{right})^2\right) & \text{if } T > T_{max} \end{cases} \tag{1b}$$

$$h(T) = h_{max}\begin{cases} \frac{D}{D+([T-T_{max}]/w_{left})^2} & \text{if } T \le T_{max} \\ \frac{D}{D+([T-T_{max}]/w_{right})^2} & \text{if } T > T_{max} \end{cases} \tag{1c}$$

$$h(T) = h_{max}\begin{cases} \frac{w_{left}}{w_{left}+[T-T_{max}]^2} & \text{if } T \le T_{max} \\ \frac{w_{right}}{w_{right}+[T-T_{max}]^2} & \text{if } T > T_{max} \end{cases} \tag{1d}$$

Table 1
The variable parameters in (1) and their “target” values; in (1a) and (1c) $D = 7.5 \times 10^{-2}$ was fixed

Formula	$h_{max}\left[\frac{J}{s\cdot m^2\cdot K}\right]$	T_{max} $[K]$	w_{left} $[K]$	w_{right} $[K]$
(1a)	0.0035	–	$w = 89.5$	–
(1b)	5700.0	680.0	260.0	80.0
(1c)	6000.0	680.0	350.0	100.0
(1d)	10000.0	680.0	3500.0	1000.0

$$\frac{\partial}{\partial r}\left(k(T(r,t))\frac{\partial T(r,t)}{\partial r}\right) + \frac{k(T(r,t))}{r}\frac{\partial T(r,t)}{\partial r} = \rho C_p(T(r,t))\frac{\partial T(r,t)}{\partial t} \tag{2}$$

Under the normal environmental pressure and the given temperature range the density of the alloy is constant $\rho = 8420$ $\left[kg\cdot m^{-3}\right]$, while the heat conductivity-temperature function $k(T)$ $\left[J\cdot s^{-1}\cdot m^{-1}\cdot K^{-1}\right]$ can be well described by a third order polynomial fitted to the tabulated data found in [2]. The same holds for the specific heat-temperature function $C_p(T)$ $\left[J\cdot kg^{-1}\cdot K^{-1}\right]$. The approximations applied are given in (3).

$$k(T) = \sum_{\ell=0}^{3} a_\ell \cdot (T/100)^\ell \quad , \quad \rho C_p(T) = \sum_{\ell=0}^{3} b_\ell \cdot (T/100)^\ell \tag{3}$$

with $a_0 = 14.398632059$, $a_1 = 1.479338274$, $a_2 = 0.0206104081$, $a_3 = -0.0006946666$, $b_0 = 3660692.67678371$, $b_1 = 466878.975781679$, $b_2 = -120161.302924857$, and $b_3 = 11898.306953622$.

Since the location of the temperature sensors in the standard ISO 9950 corresponds to $r = 0$ where (2) is singular, for developing “Finite Elements Method (FEM)” for computing the *HTC* and the cooling curves in the calculations the following *approximations* were done. For the probe of radius $R = 6.25\,[mm]$, instead of the *exact range* $[0,R]$ the practically computable range $[\Delta r, R]$ was so considered that the $[0,R]$ interval was divided into $N \in \mathbb{N}$ equally long subintervals as $\Delta r = \frac{R}{N}$, and in the role of the centre line $r = \Delta r$ was placed. For the grid points $\{r_i | i = 2,\dots,N-1\}$ the *central estimation of the gradient* was used as in (4a). In this manner it was possible to calculate the 2nd term at the LHS of (2). Again, by the application of the central differences, the calculation of the 1st term at the LHS of (2) was possible only for the points $\{r_i | i = 3,\dots,N-2\}$ in (4b). The temperature in the centreline was estimated according to (4c). This scheme allowed the estimation of $\partial T/\partial t$ for the same grid-points, and on this basis, the application of a simple Euler-integration according to the time as $T(r_i, t+\delta t) \approx T(r_i,t) + \delta t \frac{\partial T(r_i,t)}{\partial t}$.

$$\nabla T(r_i,t) \equiv \frac{\partial T(r_i,t)}{\partial r_i} \approx \frac{T(r_{i+1},t) - T(r_{i-1},t)}{r_{i+1} - r_{i-1}} \quad , \qquad i \in \{2,\dots,N-1\} \tag{4a}$$

$$\frac{\partial}{\partial r}\left(k\frac{\partial T}{\partial r}\right) \approx \frac{k(r_{i+1},t)\nabla T(r_{i+1},t) - k(r_{i+1},t)\nabla T(r_{i-1},t)}{r_{i+1} - r_{i-1}} \quad , \qquad i \in \{3,\dots,N-2\} \tag{4b}$$

$$T(r_1,t) \equiv T(r_2,t) \equiv T(r_3,t) \tag{4c}$$

The boundary conditions were set by (5)

$$-k(T(R,t))\frac{\partial T(R,t)}{\partial r} = h(t)(T(R,t) - T_q) \tag{5}$$

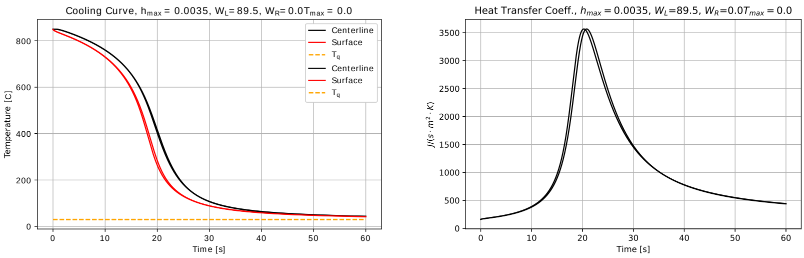
in which $T_q\,[C]$ is the temperature of the bulk quenching liquid in turbulent flow, and $h(t)\,\left[J \cdot s^{-1} \cdot m^{-1} \cdot K^{-1}\right]$ is the *heat transfer coefficient* of the boundary layer of the liquid at the surface of the probe. Besides the boundary condition equation (2) must be completed with the *initial conditions* that have to be compatible with the boundary conditions, too. This compatibility was guaranteed as follows: to solve the boundary conditions in (5) a *refreshed value* T in grid point r_{N-1} was estimated by using the 1st backward spatial derivative of the already refreshed points as in (6)

$$T(r_{N-1},t) \approx T(r_{N-2},t) + \frac{(r_{N-1} - r_{N-2})(T(r_{N-2},t) - T(r_{N-3},t))}{(r_{N-2} - r_{N-3})} \tag{6}$$

and $T(r_N,t)$ was estimated by the use of the heat transfer coefficient as in (7)

$$T(r_N,t) \approx \frac{h(T(r_{N-1},t))T_q + k(T(r_{N-1},t))/\Delta r}{h(T(r_{N-1},t)) + k(T(r_{N-1},t))/\Delta r} \tag{7}$$

To make the above construction *practically useful* the number of the grid-points N and the step-length of the Euler-integration according to the time, δt, must be determined. This question is critical due to the singularity in (2) at $r=0$. For this purpose the *"trial and error"* method was chosen: the cooling curves were calculated for Planck's model in (1a) for the pairs $\{N=20,\, \delta t=10^{-2}\,[s]\}$, and $\{N=100,\, \delta t=10^{-3}\,[s]\}$, and the graphs were plotted in the same charts in Fig. 1. It reveals that for the calculations it is sufficient to use the coarser one that considerably reduces the computational burden and time.



a) *b)*

Figure 1
The cooling curves (a) and the HTC values for Planck's model in (1a) for the pairs $\{N=20,\, \delta t=10^{-2}\,[s]\}$, and $\{N=100,\, \delta t=10^{-3}\,[s]\}$ (b)

3 The basic optimization algorithm

In the suggested approach, to represent the "target cooling curve", one of the "format functions" in (1) can be selected with the parameter values given in Table 1. Then, by applying the FEM method described in Section 2, for the points of the *discrete time-grid applied*, the cooling curve $T_c \in \mathbb{R}^L$, $L \in \mathbb{N}$ as an array can be computed. Following that the error of the difference of the arrays defined as $E \overset{def}{=} \|T_c - T_{target}\|^{\nu}$ can be computed. Finally, the scalar function $E(x) \equiv E(h_{max}, T_{max}, w_{left}, w_{right}) : \mathbb{R}^4 \mapsto \mathbb{R}$ can be minimized by varying its 4 independent arguments. Normally, in the practice using quadratic cost functions (i.e. $\nu=2$ in the Frobenius norm) happens because of rather "technical" than "mathematical" reasons. In the present simulation it was found that the choice of $\nu=0.25$ produces a "convenient" cost function. If the local optimum must be approached under constraints, Lagrange's *"Reduced Gradient Method"* can be applied for a given initial argument x_0. If there are no constraints, the Gradient Descent method generates a sequence of approximations as $x_{i+1} = x_i + \alpha_i \nabla E_i \;\;,\;\; i \in \{0,1,2,\ldots\}$ in which α_i is a small negative number if the aim is to minimize $E(x)$. If no *a priori* information is available on $\|\nabla E(x)\|$, very small α_i must be chosen. If some assumption is available for the minimum, a big step can be done in the direction of the gradient that could approach this minimum in a single step, then the same consideration can be repeated until reaching this minimum *"Newton-Raphson Algorithm"*. Since this method converges well only if the available minimum can be precisely estimated in advance, we have the freedom to choose various α_i values.

The idea of accelerating the Newton-Raphson Algorithm consists of the combination of various solutions borrowed from already known methods as follows: i) From the *Simplex Algorithm* the idea of shrinking the simplex in the vicinity of the local optimum is borrowed. The speed of the motion of the simplex roughly is proportional to its size. It can quickly approach the local minimum but its drift stops in its vicinity. After shrinking, the smaller simplex drifts with reduced speed but finds the optimum with better precision. For this purpose a *"Shrinking Parameter"* $\upsilon \in (0,1)$ is introduced as $\alpha \Leftarrow \upsilon \cdot \alpha$. ii) From the *Particle Swarm Optimization* the idea of storing the values of the already found best solution is borrowed. iii) Similar problem may happen when the shrinking operation occurs too frequently. This means the appearance of very small factors containing $\upsilon^k \to 0$ as $\mathbb{N} \ni k \to \infty$, therefore the algorithm may become very slow. In the calculations $\upsilon = 0.6$ was chosen.

The Newton-Raphson algorithm has alternatives as e.g. the *"Fixed Point Iteration"*-based methods that may work well, too. Their essence is Banach's Fixed Point Theorem [11] according to which the solution of certain numerical problems can be reformulated as finding the fixed point of a contractive map defined over a complete linear normed metric space. This fixed point subsequently can be approached by a simple iteration as follows. Consider the (generally nonlinear) real function $f : \mathbb{R}^n \mapsto \mathbb{R}^n$, $n \in \mathbb{N}$ with the mathematical problem $q^{Des} = f(q_\star)$ in which q^{Des} is a known *output* (the "response") for which we have to find the appropriate *input* $q_\star$. Dineva in [12] suggested the iteration for finding the appropriate input defined as a sequence of deformed inputs as:

$$q^{Def}(i+1) = [F(A_c \|h(i)\| + x_*) - x_*]\, e(i) + q^{Def}(i) \tag{8}$$

with the *"response error"* defined as $h(i) \overset{def}{=} f\left(q^{Def}(i)\right) - q^{Des}(i+1)$, and the vector of unit Frobenius norm $e(i) \overset{def}{=} \frac{h(i)}{\|h(i)\|}$. Here $A_c \in \mathbb{R}$ is the *adaptive control parameter*, and $F : \mathbb{R} \mapsto \mathbb{R}$ is a real differentiable function with an attractive fixed point $F(x_\star) = x_\star$. Evidently, if $h(i) = 0$ then $q^{Def}(i+1) = q^{Def}(i)$, that is the solution of the control task is the fixed point of this problem. By considering the 1st order Taylor series approximation of $F(x)$ around $x_\star$ and $f(q^{Def})$ around $q_\star$ Dineva proved that an appropriate A_C can be chosen for obtaining a convergent sequence if the real part of each eigenvalue of $\left.\frac{\partial f}{\partial q^{Def}}\right|_{\ddot{q}_\star} \in \mathbb{R}^{n\times n}$ is either positive or negative. In (8) various $F(x)$ functions can be chosen. Instead of choosing some analytical formula in [13] a special function was applied that can be realized and well configured in a program block. It transforms the vector $b \in \mathbb{R}^n$ into the vector $a \in \mathbb{R}^n$ ($\|a\| \neq \|b\|$) via so augmenting them with a physically not interpreted $(n+1)^{\text{th}}$ dimension that the augmented vectors have the same Frobenius norm. Then an orthogonal matrix is analytically computed that rotates the augmented variant of $\mathbf{b}$ into that of $\mathbf{a}$ while leaves their orthogonal subspace invariant. With an interpolation parameter $\lambda \in (0,1)$ the angle of the full rotation can be multiplied, and this "moderated rotation" can be applied. The projection of the rotated augmented vector in the original space suffers simultaneous rotation and shrinking/dilatation as it approaches vector $\mathbf{a}$. This idea evidently can be applied for $\nabla E(x)$ that can be driven to zero by fixed point iteration if in the place of the desired value $q^{Des} \equiv 0$, in the role of q^{Def} the variable x, and in the role of the realized value $f(q^{Def})$ the quantity $\nabla E(x)$ are written: $x_{n+1} = \Phi(\nabla E_n, x_n, 0)$. Roughly speaking, in the vicinity of the local minimum, by driving $\nabla E(x)$ to zero by a simple fixed point iteration can be applied. In the lack of information whether the actual point is in the basin of attraction of a local minimum or a local maximum (we wish to evade the calculation of $J := \frac{\partial^2 E}{\partial x_i \partial x_j}$), this algorithm may proceed toward a local maximum while its counterpart based on the Gradient Descent approach always moves towards the local minimum. The speed of its migration also depends on the Jacobian J. However, the behaviour of the fixed point iteration can be better and can be worse than that of its gradient descent counterpart. To tackle this problem the following procedure was applied: the Fixed Point Iteration was modified as follows: to evade too big jumps, in the goal for the step $(i+1)$ instead of 0 the reduced variant of the previous value κE_i has been written with $\kappa \in (0,1)$. In this manner a finite value slowly and cautiously can be driven toward 0. To speed up the convergence, the next point in the space of the independent variables was selected as

$$\begin{aligned} x_{n+1} &= \Phi(\nabla E_n, x_n + \mu_n(x_n - x_{n-1}), \kappa \nabla E_n) \quad , \\ \mu_n &= \frac{\omega_4 \tanh(1/(\omega_3 + \widetilde{\|J\|}))}{\tanh(1/\omega_3)} \end{aligned} \tag{9}$$

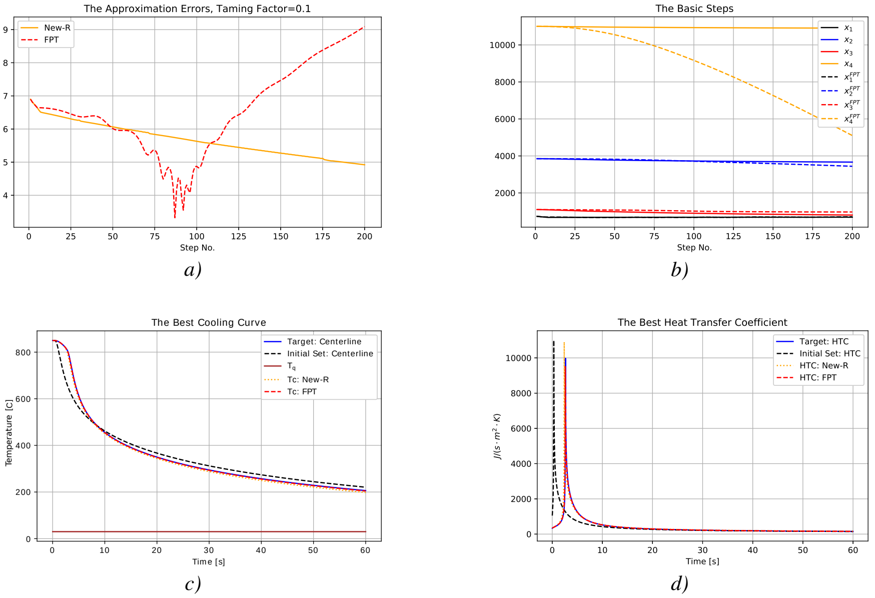
in which $\widetilde{\|J\|}$ is a roughly estimated, filtered "Jacobian". Regarding the filter, by introducing a *"smoothing and forgetting factor"* $\eta \in (0,1)$, and utilizing that $\sum_{\ell=0}^{\infty} \eta^{\ell} = 1/(1-\eta)$, any discrete time-dependent quantity $f(t_i)$ can be replaced with its filtered value $\tilde{f}(t_i) = (1-\eta)\sum_{\ell=0}^{\infty} \eta^{\ell} f(t_{i-\ell})$. This corresponds to the weighted average of the recent values of f in which the very old contributions are gradually forgotten due to their small weighting coefficients. In the calculations $\eta = 0.9$ was applied. Evidently, if $\widetilde{\|J\|} \ll \omega_3$, then $\mu_n \approx \omega_4$, that means that for very slow process the "accelerator" factor has some limit. For $\widetilde{\|J\|} \to \infty$ $\mu_n \to 0$, i.e. if the process is fast enough, practically no process acceleration happens. The simple approximation in (10) was applied.

$$\widetilde{\|J\|}_n = \frac{\widetilde{\|\nabla E_n - \nabla E_{n-1}\|}}{\widetilde{\|x_n - x_{n-1}\|} + \omega_2} \tag{10}$$

in which $\omega_2 > 0$ has the role of evading division by zero. The computation of (10) evidently means far less burden than the numerical estimation of the real Jacobian. However, *it does not contain information on the direction of the drift of* x, so it can be quite unreliable, and may not result in monotonic decrease of E. However, it sooner may produce better values than the Newton-Raphson algorithm, and this better value can be stored and used even if the estimation based on (10) later diverges. In the computations $\kappa = 0.25$, $\omega_4 = 1.0$, $\omega_3 = 3.0$, $\omega_2 = 10^{-6}$, $\tilde{R} = 10^2$, and $\lambda = 0.1$ were used. For 5 steps the Newton-Raphson method run to create the "antecedents" for the fixed point iteration, then the computations turned to it. In Section 4 typical numerical results are provided.

4 Numerical computations

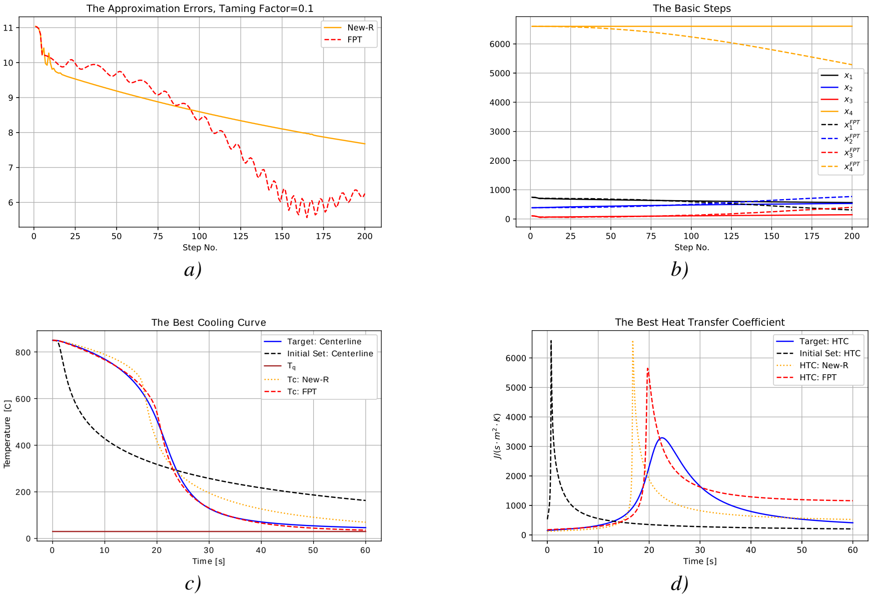
The time need of the computations depends on the hardware and software applied. The computations were made on a **Dell inspiron 15R laptop** operated by the central processor Intel® Core™ i5-3337U CPU @ 1.80GHz × 4 under Ubuntu ver. 13.04 operating system without using graphical acceleration. The sequential program was written in Julia Version 1.0.3 (2018-12-18). This program language is developed at the MIT, it is a free software, it is very similar to the MATLAB, but it runs almost as fast as a C code [14]. It provides its users with a standard macro that measures the time of computing the value of a function. It was experimentally found that for the calculation of E and ∇E approximately $0.9-1.2/s$ was used, the abstract rotations were calculated during $45-90/\mu s$. During the research 8 different scenarios were investigated. In each of them a Newton-Raphson algorithm and a FPI-based solution were compared according to the already given parameter settings. In the first group the cases with possible exactly 0 approximation error (each "target" was created by the same format function using different starting parameters) were considered. It was found that the occurrence of the exponential function in the definitions (1a) and (1b) needed quite slow and cautious approximation procedure, therefore their use was less successful. However, the present settings was successful for the approximation of the functions in (1c) and (1d) that do not contain the exponential function. In Fig. 2 the target in (1d) was approximated with the initial element (1d). In this case the FPI-based approach very early provided the best solution.



a) *b)* *c)* *d)*

Figure 2
The cost function (Error) (a), the tuned variables $x = [T_{max}, w_{left}, w_{right}, h_{max}]$ (b), the cooling curve (c), the heat transfer coefficient HTC (d)

In the following runs variants in (1c) and (1d) as initial HTC distributions were used to approximate targets created by some different distribution. When the target (1c) was approximated by the form function (1d) the Newton-Raphson method provided the better approximation. The reversed problem, i.e. when the target (1d) was approximated by the form function (1c) the FPI-based solution was better but the Newton-Raphson-based approach resulted in quite good approximation, too. Figure 3 corresponds to the approximation of (1a) by the function (1c). In this case the FPI-based approach yielded a surprisingly good result.



a) *b)* *c)* *d)*

Figure 3
The cost function (Error) (a), the tuned variables $x = [T_{max}, w_{left}, w_{right}, h_{max}]$ (b), the cooling curve (c), the heat transfer coefficient HTC (d)

5 Conclusion

In this paper a simple computational method was suggested for numerically tackling the inverse heat conduction problem that has great significance in metallurgy. By utilizing the geometric and thermodynamic simplifications offered by the standard ISO 9950 a single dimensional space r and the time t variables were considered by using the thermal data of the alloy Inconel 600. The finite elements approach, dealing with the singularity in the centre of the polar coordinates, the boundary conditions and the initial conditions that must be compatible with the boundary conditions were discussed in details. For setting the appropriate grid-points in space and time the method of *"trial and error"* was applied. A novel solution was outlined that works with an "entity" that evolves according to the Newton-Raphson Algorithm with step-by-step reduction of its step length, and also computes the propagation of an associated "partner entity" that evolves according to a fixed point iteration. While it can be taken for granted that the original “entity” converges to a local optima but its speed of convergence becomes very slow, the “partner entity” may have very fast convergence but it may diverge, too. Numerical simulations were elaborated for the combination of four typical “format functions” that can capture the essential asymmetry in the $HTC(T)$ function, and numerically can approach the measured data taken from the literature. Regarding the computational burden, the requirements of both approaches are comparable. The computations were realized by a freely available program language (Julia) that almost as efficiently utilizes the available hardware as a C code. No any special FEM software was applied. While the computation of the gradient of the cost function needed the time between $0.9\,[s]$ and $1.2\,[s]$, the time need of the calculation of the abstract rotation applied by the FPI-based solution was between $45\,[\mu s]$ and $85[\mu s]$, i.e. it was negligible in comparison with the computational needs of the error gradient.

The preliminary results indicate that there are more or less plausible possibilities for further speeding up the recommended approach as follows: a) instead of the complete array taken from the time grid under consideration more sparse samples can be used for the calculation of the error function that has to be minimized; b) instead applying more dense points in the vicinity of the important “segments” of the target cooling curve, enhanced weighting of the contribution of these points in the error function may be practical, too; c) due to the special shape of a particular cooling curve the success of approximation strongly depends on the starting parameters of the approximation; it seems to be practical to simultaneously run several “associated couples” pairs , and finally select the best result. Since the considered approximations needed very limited running time on a “common” hardware, it can be concluded that it will be expedient to conduct further modelling investigations. The present approximation is so simple that it seems to be useful in the education, too.

Acknowledgement

The financial support of this work by Hungarian-Mexican bilateral project (2019-2.1.11-TÉT-2019-00007) is acknowledged.

References

[1] ISO9950, *ISO 9950: Industrial Quenching Oils – Determination of Cooling Characteristics Nickel-Alloy Probe Test Method 1995(E).* ISO, Switzerland, 1995.

[2] J. Clark and R. Tye, “Thermophysical properties reference data for some key engineering alloys,” *High Temperatures – High Pressures*, vol. 35–36, pp. 1–14, 2003-2004.

[3] D. Landek, J. Župan, and T. Filetin, “A prediction of quenching parameters using inverse analysis,” *Materials Performance and Characterization*, vol. 3, no. 2, pp. 229–241, 2014.

[4] S. Szénási and I. Felde, “Configuring genetic algorithm to solve the inverse heat conduction problem,” *ACTA POLYTECHNICA HUNGARICA*, vol. 14, no. 6, pp. 133–152, 2017.

[5] L. Zadeh, “Fuzzy sets,” *Information and Control*, vol. 8, pp. 338–353, 1965.

[6] J. Dombi, “A general class of fuzzy operators, the de Morgan class of fuzzy operators and fuzziness measures induced by fuzzy operators,” *Fuzzy Sets and Systems*, vol. 8, pp. 197–216, 1982.

[7] J. Dombi, T. Jónás, and Z. Tóth, “The epsilon probability distribution and its application in reliability theory,” *Acta Polytechnica Hungarica*, vol. 15, no. 1, pp. 197–216, 2018.

[8] J. Tenreiro Machado and V. Kiryakova, “The chronicles of fractional calculus,” *Fract. Calc. Appl. Anal.*, vol. 20, no. 2, pp. 307–336, 2017.

[9] I. Felde, “Liquid quenchant database: determination of heat transfer coefficient during quenching,” *Int. J. Microstructure and Materials Properties*, vol. 11, no. 3/4, pp. 277–287, 2016.

[10] R. Mastrullo, A. Mauro, A. Rosato, and G. Vanoli, “Comparison of R744 and R134a heat transfer coefficients during flow boiling in a horizontal circular smooth tube,” *Proc. of the International Conference on Renewable Energies and Power Quality (ICREPQ’09), 15th to 17th April, 2009, Valencia, Spain*, vol. 1, no. 7, pp. 577–581, 2009.

[11] S. Banach, “Sur les opérations dans les ensembles abstraits et leur application aux équations intégrales (About the Operations in the Abstract Sets and Their Application to Integral Equations),” *Fund. Math.*, vol. 3, pp. 133–181, 1922.

[12] A. Dineva, J. Tar, and A. Várkonyi-Kóczy, “Novel generation of Fixed Point Transformation for the adaptive control of a nonlinear neuron model,” *In proc. of the IEEE International Conference on Systems, Man, and Cybernetics, October 10-13, 2015, Hong Kong (SMC 2015)*, pp. 987–992, 2015.

[13] B. Csanádi, P. Galambos, J. Tar, G. Györök, and A. Serester, “A novel, abstract rotation-based fixed point transformation in adaptive control,” *In the Proc. of the 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC2018), October 7-10, 2018, Miyazaki, Japan*, pp. 2577–2582, 2018.

[14] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, “Julia,” *https://julialang.org*, 2019.

[15] S. Szénási, “Solving the inverse heat conduction problem using NVLink capable power architecture,” *PeerJ Computer Science*, vol. 3, p. e138, nov 2017.