HUNGARIAN JOURNAL OF
INDUSTRY AND CHEMISTRY
Vol. 47(1) pp. 1–83 (2019)
hjic.mk.uni-pannon.hu

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

Contents

HUNGARIAN JOURNAL OF
INDUSTRY AND CHEMISTRY
Vol. 47(1) pp. 1–1 (2019)
hjic.mk.uni-pannon.hu
DOI: 10.33927/hjic-2019-01

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

EDITORIAL PREFACE TO THE SPECIAL ISSUE DEDICATED TO THE 15TH INTERNATIONAL PHD WORKSHOP ON SYSTEMS AND CONTROL

ATTILA MAGYAR[1]

[1]Department of Electrical Engineering and Information Systems, University of Pannonia, P. O. Box 158, Veszprém, H-8201, HUNGARY

Systems and control theory is a scientific area that is constantly developing and a dominant driving force behind key industries and fields of engineering, e.g. process engineering, automotive engineering, bioengineering, the energy sector, etc. The goal of this issue is to provide an overview of the actual research topics pursued by some selected PhD students. The papers presented here were chosen from among the contributions at the 15th International PhD Workshop on Systems and Control held August 30-31, 2018. The objective of this workshop was to establish an international forum for discussion between young researchers and engineers from the industry and related research fields. The meeting provided opportunities for the participants to present and discuss their latest results and up-to-date applications in systems and control.

This issue represents the entire spectrum of systems and control engineering as follows:

- process modeling and analysis; control (traditional, intelligent, adaptive, etc.)
- system identification and signal processing
- electrical transmission systems, smart grids
- bioengineering
- traffic control
- reaction kinetic networks
- artificial intelligence
- soft computing (neural, genetic, fuzzy algorithms, etc.)
- software issues (parallel computing, distributed and network computing, data visualization)
- decision making (decision support, data mining)
- applications

The organizers are grateful to the authors for their contributions. The tradition of the International PhD Workshop on Systems and Control continues.

You are welcome to participate at the 16th International PhD Workshop on Systems and Control in Veszprém, 2020.

Attila Magyar
Guest Editor of the Issue

HUNGARIAN JOURNAL OF
INDUSTRY AND CHEMISTRY
Vol. 47(1) pp. 3–9 (2019)
hjic.mk.uni-pannon.hu
DOI: 10.33927/hjic-2019-02

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

NON-TECHNICAL LOSS DIAGNOSIS IN ELECTRICAL NETWORKS WITH A RADIAL LAYOUT

ANNA I. PÓZNA *[1], ATTILA FODOR[1], AND KATALIN M. HANGOS[2]

[1]Department of Electrical Engineering and Information Systems, University of Pannonia, P. O. Box 158, Veszprém, 8201, HUNGARY
[2]Institute for Computer Science and Control, Hungarian Academy of Sciences, P. O. Box 63, Budapest, 1518, HUNGARY

A network-oriented non-technical loss detection and localization methodology is presented in this paper. The basic idea behind the proposed methodology is the deviation of the measured voltages from their nominal values. The operation of the algorithm was investigated by simulation experiments using an (IEEE) European Low Voltage Test Feeder benchmark network. The simulation results show that the proposed method is able to detect and localize multiple occurrences of non-technical losses caused by fraudulent meters.

Keywords: power network, diagnosis, non-technical loss, load-flow

1. Introduction

The demand for electrical energy is continuously increasing on a global scale. Electrical energy is generated by power plants or renewable energy sources and is transmitted from the source of generation to consumers through distribution stations and power lines. During transmission, technical losses, that originate from dissipation in conductors, transmission lines and substation transformers as well as magnetic losses in transformers, reduce the efficiency of power delivery. The proportion of technical losses is about 20 % of the total energy transmitted.

Besides the technical losses, non-technical losses (NTLs) may be present as well. These are unnecessary losses which are not expected and cannot be anticipated. The NTLs are usually related to energy theft and fraudulent consumption behavior. Energy theft has been a widespread and major issue for many years and various techniques of energy theft are present from unregistered users to hacking meters [1]. Following this unwanted phenomena, several methods of non-technical loss detection have appeared in the literature. It can be stated that no golden rule exists for detecting energy theft, rather, several different approaches are available [2]. Papers [3] and [4] provide very good reviews on the most frequently used methods in this field.

The majority of solutions available in the literature are based on the analysis of consumption data using some statistical or machine-learning methods, for example, the authors of [5] used a linear regression-based procedure that not only detects energy theft but defects of smart meters as well. A probabilistic neural network-based classification approach is presented in [6] where the Levenberg-Marquardt method is used for training the network. A support-vector machine-based solution is given in [7], where a parallel computer architecture was proposed in order to enhance computation.

Another approach to non-technical loss detection is based on network topology, such network-oriented methods use readings from grid sensors and smart meters. In [8], the authors proposed state estimation with a Kalman filter to identify currents and biases in a microgrid network. The currents and biases are estimated separately using two different filters. If the estimated bias of a customer exceeds the predefined threshold, then this user has committed fraud. The authors of [9] suggest a probabilistic power flow approach to NTL detection. The output of the algorithm is a probability distribution of NTLs in the subnetwork.

Besides the above classes, other methods of localizing illegal electricity usage exist, e.g. in [10] a power lines inspection robot was applied to find NTLs.

The approach followed in the present work belongs to the network-oriented class and it is based on analyzing the differences between the measured and nominal voltages. The uncertainty in the model parameters together with the measurement uncertainties are taken into account to ensure the approach is applicable to real-world cases.

The structure of the paper is as follows. The basic notions and problems are introduced in Section 2. Section 3

*Correspondence: pozna.anna@virt.uni-pannon.hu

contains the main contribution of the paper and presents the proposed novel diagnostic method in detail. Afterwards, the proposed method is subject to simulation-based analysis in Section 4.

2. Problem statement

2.1 Non-technical losses

According to a source Non-Technical Losses (NTLs) can be classified as follows [11]:

Before meter: illegal tapping of distribution lines or feeders

Meter: inaccurate power readings due to a meter being faulty (e.g. changes to the calibration), reversed, disconnected or bypassed

Billing: non-payment of electricity bills, inaccurate billing, faulty operation of the billing system, a cyber attack against the billing system, etc.

NTLs caused by companies in the US were estimated to cost between 0.5 % and 3.5 % of annual gross revenues. The NTLs may account for more than 15 % of the generated power in some countries [2] and [12].

In the remaining part of the paper only the following type of non-technical loss is examined that belongs to the first class above:

- fraud committed by tampering with the meter (reversing or hacking the hardware or software, or calibration of the meter),
- bypassing the meter.

2.2 Detection and localization

In general, two main parts of the diagnostic procedure exist: fault detection and fault identification. The aim of detection is to decide if any fault has occurred in the system. The exact type and localization of the occurred fault is determined in the identification phase. In the case of non-technical loss diagnosis, detection and localization are defined as follows:

Detection of non-technical loss means that the loss is acknowledged.

Localization means that the fraudulent user is identified if the illegal consumption of power occurs at the meter. In the case of multiple NTLs, every fraudulent meter is to be identified by the method.

Illegal load is a load that originates from a fraudulent meter, i.e. the consumer uses illegal power.

2.3 Basic assumptions

During development of the non-technical loss detection and localization method, the following assumptions were made:

- The electrical network is represented by its static linear model. The known parameters of the model are the resistances of the transmission lines, the current and voltage of the feeding point (transformer), and the currents of the loads.
- The structure of the electrical network is radial (for more details see Section 3.2).
- Every load has a smart meter that measures the current, voltage and power consumption of the load.
- At least one illegal load may be present in the network. If more than one illegal load exists then each is located in a different part of the network (see Section 3.2 for more details).

2.4 Uncertainties and measurement errors

The main uncertainties that affect the voltages and currents in a public electrical grid can be classified as follows:

- *Uncertainties about the parameters of the transmission lines*
  The resistance of the transmission lines is the main source of uncertainty. This resistance can be computed from its diameter, curvature and length, which are functions of temperature. The function is approximately linear in the domain -50 °C to 100 °C:
  $$\rho = \rho_0(1 + \alpha(T - T_0)), \tag{1}$$
  where ρ and T stand for the values of resistance and temperature, respectively, while ρ_0 and T_0 denote the nominal resistance and temperature, respectively, and finally α represents the temperature modulus. The main uncertainty over the line is the small difference between the planned and installed transmission line. The losses, which are justifiable given these technical reasons, contribute to approximately 2–3 %.
- *Measurement errors*
  The presence of smart meters in our electrical network is assumed. The precision of smart meters varies from ± 0.2 % to ± 2 % depending on their precision and the percentage of nominal power, see International Standards IEC 62051, IEC 62052-11, IEC 62052-21 and IEC 62052-31.

In this paper the effects of the harmonic currents [13] are not investigated, rather, it is assumed that the voltages and currents are sinusoidal. The effects of asymmetrical loads [14] are also not investigated because a single-phase grid is assumed. A three-phase grid can be assembled from three single-phase grids with one N line.

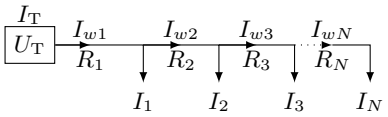


Figure 1: Single-feeder layout

3. Diagnostic method

The proposed diagnostic method uses the topology and electrical parameters of the network together with the nominal and measured voltages as well as current values to detect and localize the illegal loads. The nominal voltage and current values are generated from the topology and the parameters of the network determined by solving its linear time-invariant model using the known feeder voltage and current values [15].

Input data It is assumed that the model of the network is present together with the nominal values of its elements. Current and voltage readings with measurement errors are available from smart meters of all loads.

Detection The proposed diagnostic method is based on analyzing the difference between the measured and nominal voltages. It is assumed that the currents measured are the nominal currents of the network. The presence of an illegal load is detected if a difference between the sum of the measured currents $\left(\tilde{I}_i, i = 1, \ldots, N\right)$ and the measured current of the transformer (I_T) exists.

$$\sum_{i=1}^{N} \tilde{I}_i - I_T > \varepsilon \tag{2}$$

Localization The localization method starts with the simulation of the network assuming the nominal current values. During the simulation, the voltages of the loads are computed. The simulated voltages are considered to be the nominal voltages. Subsequently the nominal voltages are compared to the measured voltages. Localization is based on the evaluation of deviations in voltage from the nominal values. The measurement error is taken into account in such a way that only deviations that fall outside the maximum measurement error are considered during the diagnosis.

3.1 Single-feeder layout

The single-feeder layout is the simplest topology of electrical networks. It contains a single feeding point with several loads connected to it along a transmission line (Fig. 1). The difference between the nominal and measured voltage levels is computed as:

$$\Delta U_i = \tilde{U}_i - U_i,\ i = 1, ..., N, \tag{3}$$

where $\tilde{U}_i$ and U_i are the measured and nominal voltages of the ith load, respectively. Larger drops in voltage than the nominal ones are caused by illegal loads. Therefore,

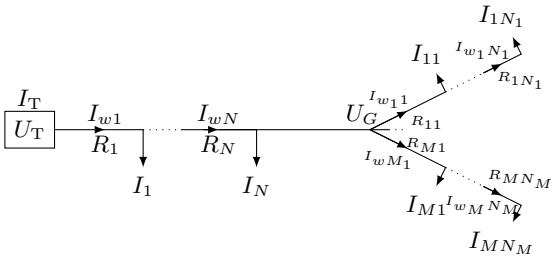


Figure 2: Radial feeder layout

the value of ΔU_i is negative. If the illegal load is in the kth load, differences in voltage of subsequent loads are equal to the kth load. This difference is proportional to the magnitude of the illegal current and the resistance of the transmission line:

$$\Delta U_i = \begin{cases} I_{\text{ill}} \sum_{j=1}^{i} R_j, & \text{if } i < k \\ I_{\text{ill}} \sum_{j=1}^{k} R_j, & \text{if } i \geq k \end{cases} \tag{4}$$

If, by starting from a certain load, the voltage differences of the loads are equal, then the load in question consumes power illegally. In practice two differences in voltage are considered to be equal if the difference between them is less than the measurement error. Therefore, localization of the illegal load is quite straightforward in a single-feeder layout: the load in question needs to be identified and from this the differences in voltage start to become equal.

This method can be generalised if more than one illegal load in a single-feeder network is present. In this case, sections in the sequence of voltage differences where consecutive voltage differences are equal exist. The illegal loads are located at the start of these sections.

3.2 Radial layout

The radial-feeder layout is commonly used in low-voltage networks. The general structure of a radial-feeder network can be seen in Fig. 2. This type of network can be decomposed into single-feeder subnetworks by identifying and cutting off the branches (for further details see [16]).

The loads are considered to be part of the same subnetwork if they are connected to the same bus. After decomposition, a set of disjoint single-feeder networks is formed which can be processed in parallel.

From a diagnostic point of view, two types of subnetworks should be distinguished. The first type is when the subnetwork contains more than one load (referred to as *multiple load subnetwork* hereinafter). The structure of this subnetwork is similar to the single-feeder layout (Fig. 1). The second type of subnetworks is the special case when the subnetwork contains only one load (referred to as *single load subnetwork* hereinafter).

One illegal load in a multiple load subnetwork If only one illegal load is present in the whole network and it is

located in a subnetwork with several loads, then it can be clearly localized using the diagnostic method described in Section 3.1.

One illegal load in a single load subnetwork In the case when the subnetwork contains only one load, the voltage difference cannot be compared to any other voltage difference within this subnetwork. Therefore, the diagnostic procedure in Section 3.1 cannot be used. In this case, the voltage deviations need to be analyzed globally. The increased consumption causes the biggest deviation in the voltage at the location of the illegal load. In this case, the diagnostic procedure to determine the minimum voltage difference is identical.

More illegal loads in multiple load subnetworks If more than one illegal load is present and are located in different multiple load subnetworks, then their locations can be determined independently of each other.

More illegal loads in single load subnetworks In this case, local minima in the voltage differences indicate the locations of the illegal loads. The local minima are determined in such a way that the voltage differences of the single loads can be compared to the voltage differences of their two nearest neighbors. If the voltage difference of the single load is smaller than that of its neighbors, then an illegal load is present at the single load.

3.3 Diagnostic algorithm

The four aforementioned cases in the radial layout can be merged into one algorithm, the flowchart of which can be seen in Fig. 3. During the diagnosis the illegal loads identified are collected in a set called NTL. At the beginning of the diagnostic algorithm, the set NTL is empty. The diagnostic algorithm searches for illegal loads in different parts of the network. The algorithm consists of three main parts: searching in multiple load subnetworks, searching for one illegal load in single load subnetworks, and searching for more illegal loads in single load subnetworks.

Detection

The inputs of the combined algorithm are the measured currents and voltages as well as the network structure. First, to detect the illegal consumption, the inequality in Eq. 2 is checked. If the difference between the sum of the measured currents and the transformer current exceeds a predefined threshold, then illegal consumption occurs in the network, otherwise the network operates normally. If illegal loads are detected, then the algorithm tries to determine their locations. To do this, the network is simulated using the measured current to obtain the nominal voltage values.

Isolation

NTLs in multiple load subnetworks This algorithm starts to search for illegal loads in the multiple load subnetworks using the method described in Section 3.1. The illegal loads identified are added to the set NTL. In this step, all of the illegal loads in the multiple load subnetworks are localized.

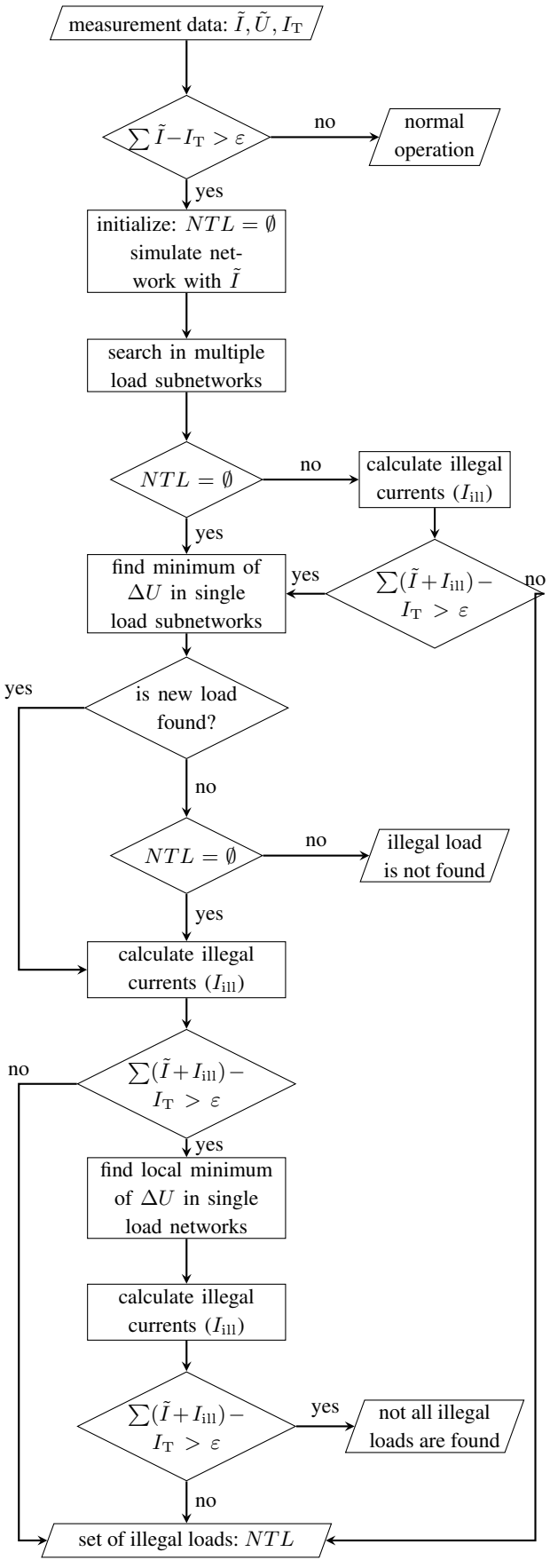


Figure 3: Flowchart of the diagnostic procedure

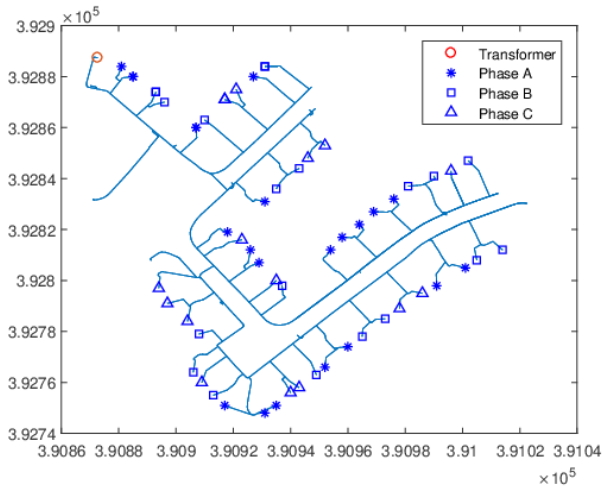


Figure 4: Structure of the IEEE 2015 European Low Voltage Test Feeder network

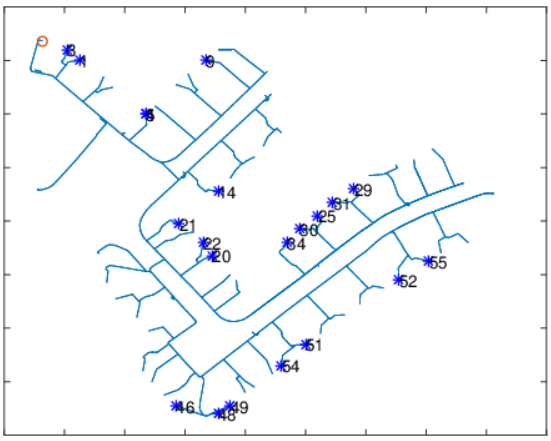


Figure 5: Loads in Phase A

Before proceeding, the algorithm checks if any illegal loads have been identified in the multiple load subnetworks. If the set NTL is empty, then no illegal loads were found, therefore, they should be located in the single load subnetworks. If at least one illegal load is identified in the multiple load subnetworks, then their illegal currents are calculated using Eq. 4. The currents of the illegal loads are substituted by the calculated currents and the inequality of Eq. 2 is checked. If the inequality is false, then all of the illegal loads are localized and the algorithm stops. If the inequality is true, then at least one illegal load is still present in the single load subnetworks.

NTL in a single load subnetwork To identify the illegal load in single load subnetworks, the algorithm searches for the minima of the voltage differences. The illegal load possesses the minimum voltage difference.

If during this step no new illegal loads are identified, then two scenarios are possible: *(i)* if set NTL is still empty, then the illegal load has not been found and the algorithm proceeds to the second search in the single load networks. *(ii)* If the set NTL is not empty, then the illegal currents are calculated and the inequality of currents checked. If no significant difference is present, then all illegal loads have been identified and the algorithm stops. If the inequality is true, then a second search of the single load networks is performed.

More NTLs in single load subnetworks During this step, the voltage difference of the single loads is compared to the voltage difference of their two nearest (left and right) neighbors. If the voltage difference of the single load is minimal between the three differences, then the single load is an illegal load.

After this search the inequality of the currents is verified again. If a difference is still present, then not all the illegal loads have been identified, otherwise all of the illegal loads have been found and the algorithm stops.

4. Case study

The diagnostic algorithm was tested on the IEEE 2015 European Low Voltage Test Feeder [17] which is a benchmark provided by the Power System Analysis, Computing & Economics (PSACE) Committee. It is a three-phase radial distribution feeder with one feeding point. The network contains 1 transformer, 55 loads and 905 lines. The structure of the network can be seen in Fig. 4.

The algorithm was tested by only taking into consideration one phase (namely Phase A) of the system. 21 loads are present in this phase which are displayed in Fig. 5 along with their identifiers. At the time of the test, the minimum and maximum power consumed by these loads was 35 W and 64.9 W, respectively. During the diagnosis, the measurement error was set at 0.2 % of the measured currents. The network belonging to Phase A can be decomposed into 13 subnetworks, 5 of which are single load networks and 8 are multiple load networks consisting of 2 loads each.

The decomposition and diagnostic algorithm was implemented in MATLAB. The simulation of the network was also performed in MATLAB [18] using the method of nodal potentials.

4.1 Case 1: One illegal load

At first, it is assumed that only one illegal load is present in the network, more specifically in a single or multiple load subnetwork. Let us consider the subnetwork that contains the loads No. 25 and 30. Load No. 25 increases its consumption by 80 % of its nominal value but deceives the current meter, therefore, the registered current value is not suspicious. The network is simulated using the nominal current values.

The difference between the nominal and measured voltages is presented in Fig. 6. It can be seen that the voltage differences are equal in the multiple load subnetworks, e.g. loads 1 and 3 as well as 20 and 22, except for the subnetwork of loads 25 and 30. Since the absolute voltage deviation at load No. 25 exceeds that at load No. 30, the illegal consumption is located at load No. 25.

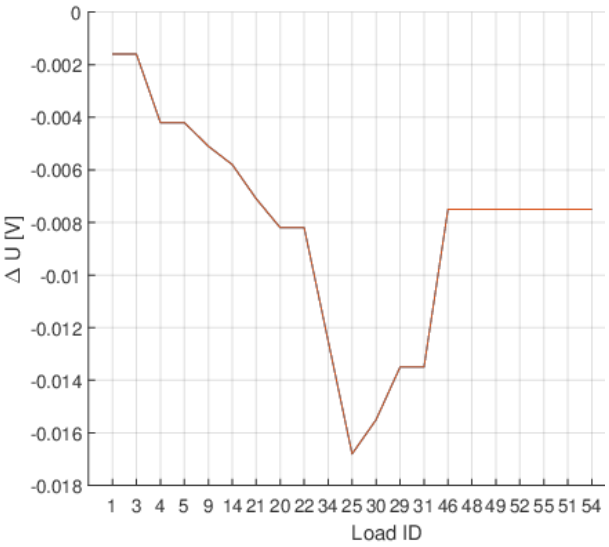


Figure 6: Voltage differences of the loads in Case 1

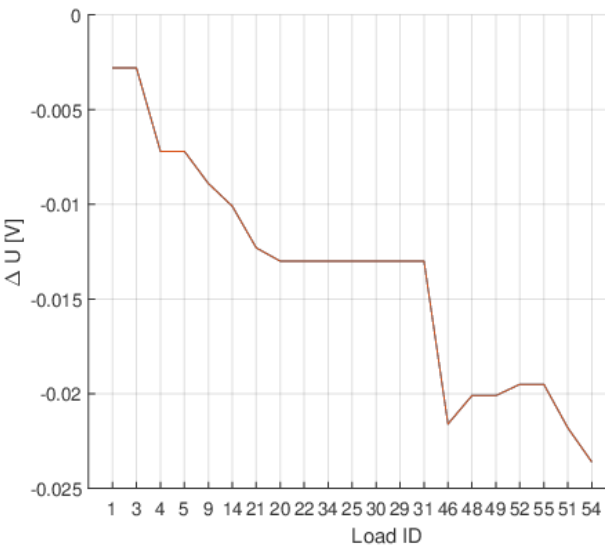


Figure 7: Voltage differences between the loads in Case 2

4.2 *Case 2: More illegal loads in different subnetworks*

In the second case, it is assumed that two illegal loads are present in different subnetworks. The first illegal load is No. 46 which is located in a single load subnetwork. The power consumption of load No. 46 is 50 % more than its nominal value. The second illegal load is load No. 54 which is located in a multiple load subnetwork along with load No. 51. The consumption of load No. 54 is increased by 80 % of its nominal value.

The difference between the simulated and measured voltages can be seen in Fig. 7. At first the diagnostic algorithm searches for illegal loads in the multiple load subnetworks. In the subnetwork of loads No. 51 and 54, the voltage deviation of load No. 54 exceeds the voltage deviation of load No. 51, therefore, load No. 54 is identified as an illegal load. The illegal current is calculated using Eq. 4. In the other multiple load subnetworks, the voltage differences of the loads within a subnetwork are equal, therefore, it can be stated that no illegal loads are present in these subnetworks.

The algorithm checks if some remaining illegal currents are present thereafter. A difference between the current of the transformer and the sum of the measured currents is still present which is indicative of at least one illegal load that is yet to be identified. These illegal loads should be located in single load networks. The minimum of the voltage differences is located at load No. 46, therefore, it is an illegal load. After calculating the illegal current of load No. 46 and checking the inequality in Eq. 2, it can be stated that no additional illegal loads are present in the network.

5. Conclusions and future work

A novel diagnostic method for detecting and locating illegal loads in electrical radial networks is proposed in this paper that utilizes the topology of the network and is capable of taking the uncertainties and measurement errors into account.

A preprocessing step decomposes the radial layout of single-feeder subnetworks with single or multiple loads, and the method is capable of locating the illegal load(s) in the subnetworks in parallel. The proposed method can detect and locate multiple independent illegal loads under certain conditions.

Future work will include the extension of the diagnostic methods to general, not necessarily radial, topology and to develop the computational model of a network to handle the uncertainties related to network parameters (resistances).

Furthermore, the effect of the uncertainties and measurement errors on the diagnostic accuracy should also be investigated.

Acknowledgement

We acknowledge the financial support of Széchenyi 2020 under the EFOP-3.6.1-16-2016-00015. We acknowledge the financial support of Széchenyi 2020 under the GINOP-2.2.1-15-2017-00038. This research is partially supported by the National Research, Development and Innovation Office - NKFIH through grant No. 115694.

REFERENCES

[1] Ahmad, T.; Chen, H.; Wang, J.; Guo, Y.: Review of various modeling techniques for the detection of electricity theft in smart grid environment, *Renew. Sust. Energ. Rev.*, 2018 **82**, 2916 – 2933, DOI: 10.1016/j.rser.2017.10.040

[2] Smith, T.B.: Electricity theft: a comparative analysis, *Energy Policy*, 2004 **32**(18), 2067 – 2076, DOI: 10.1016/S0301-4215(03)00182-4

[3] Messinis, G.M.; Hatziargyriou, N.D.: Review of non-technical loss detection methods, *Electr. Pow. Syst. Res.*, 2018 **158**, 250 – 266, DOI: 10.1016/j.epsr.2018.01.005

[4] Viegas, J.L.; Esteves, P.R.; Melício, R.; Mendes, V.; Vieira, S.M.: Solutions for detection of non-technical losses in the electricity grid: A review, *Renew. Sust. Ener. Rev.*, 2017 **80**, 1256 – 1268, DOI: 10.1016/j.rser.2017.05.193

[5] Yip, S.C.; Wong, K.; Hew, W.P.; Gan, M.T.; Phan, R.C.W.; Tan, S.W.: Detection of energy theft and defective smart meters in smart grids using linear regression, *Int. J. Elec. Power*, 2017 **91**, 230 – 240, DOI: 10.1016/j.ijepes.2017.04.005

[6] Ghasemi, A.A.; Gitizadeh, M.: Detection of illegal consumers using pattern classification approach combined with Levenberg-Marquardt method in smart grid, *Int. J. Elec. Power*, 2018 **99**, 363 – 375, DOI: 10.1016/j.ijepes.2018.01.036

[7] Depuru, S.S.S.R.; Wang, L.; Devabhaktuni, V.; Green, R.C.: High performance computing for detection of electricity theft, *Int. J. Elec. Power*, 2013 **47**, 21 – 30, DOI: 10.1016/j.ijepes.2012.10.031

[8] Salinas, S.A.; Li, P.: Privacy-preserving energy theft detection in microgrids: A state estimation approach, *IEEE T. Power Syst.*, 2016 **31**(2), 883–894, DOI: 10.1109/TPWRS.2015.2406311

[9] Neto, E.A.A.; Coelho, J.: Probabilistic methodology for Technical and Non-Technical Losses estimation in distribution system, *Electr. Pow. Syst. Res.*, 2013 **97**, 93–99, DOI: 10.1016/j.epsr.2012.12.008

[10] Oyun-Erdene, M.; Byambasuren, B.E.; Matson, E.T.; Kim, D.: Detection and localization of illegal electricity usage in power distribution line, *Multimed. Tools Appl.*, 2016 **75**(9), 4997–5012, DOI: 10.1007/s11042-014-2022-2

[11] Ahmad, T.: Non-technical loss analysis and prevention using smart meters, *Renew. Sust. Energ. Rev.*, 2017 **72**, 573–589, DOI: 10.1016/j.rser.2017.01.100

[12] Nagi, J.; Mohammad, A.; Yap, K.; Tiong, S.; Ahmed, S.: Non-technical loss analysis for detection of electricity theft using support vector machines, in Power and Energy Conference, 2008. PECon 2008. IEEE 2nd International (IEEE), 907–912, DOI: 10.1109/PECON.2008.4762604

[13] Görbe, P.; Magyar, A.; Hangos, K.M.: THD reduction with grid synchronized inverter's power injection of renewable sources, in Power Electronics Electrical Drives Automation and Motion (SPEEDAM), 2010 International Symposium on (IEEE), 1381–1386, DOI: 10.1109/SPEEDAM.2010.5545079

[14] Neukirchner, L.; Görbe, P.; Magyar, A.: Voltage unbalance reduction in the domestic distribution area using asymmetric inverters, *J. Clean. Prod.*, 2017 **142**, 1710–1720, DOI: 10.1016/j.jclepro.2016.11.119

[15] Greber, M.: Non-technical loss detection in electrical distribution network (in Hungarian), Tech. Rep., Faculty of Information Technology, University of Pannonia, 2018

[16] Pózna, A.; Fodor, A.; Gerzson, M.; Hangos, K.: Colored Petri net model of electrical networks for diagnostic purposes, *IFAC-PapersOnLine*, 2018 **51**(2), 260–265, DOI: 10.1016/j.ifacol.2018.03.045

[17] Schneider, K.; Mather, B.; Pal, B.C.; Ten, C.W.; Shirek, G.; Zhu, H.; Fuller, J.; Pereira, J.L.R.; Ochoa, L.; De Araujo, L.; et al.: Analytic Considerations and Design Basis for the IEEE Distribution Test Feeders, *IEEE T. Power Syst.*, 2018 **33**(3), 3181–3188, DOI: 10.1109/TPWRS.2017.2760011

[18] The Mathworks, Inc., Natick, Massachusetts: MATLAB version 9.0.0.341360 (R2016a), 2016

HUNGARIAN JOURNAL OF
INDUSTRY AND CHEMISTRY
Vol. 47(1) pp. 11–16 (2019)
hjic.mk.uni-pannon.hu
DOI: 10.33927/hjic-2019-03

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

EVALUATION OF RESOURCE OPTIMIZATION BASED ON QUANTUM SEARCH

SARA EL GAILY *[1]

[1]Department of Networked Systems and Services, Budapest University of Technology and Economics, Műegyetem rkp. 3-9, Budapest, 1111, HUNGARY

Quantum computing and communications try to identify more efficient solutions to the most challenging classical problems such as optimization, secure information transfer, etc. This paper will describe a new quantum method for the distribution of resources in computing platforms that consist of a large number of computing units. Furthermore, a simulation environment was developed and the performance of the new method compared to a classical reference strategy will be demonstrated. Moreover, it will be proven that the proposed solution tackles the problems of computational complexity, computing units that are time-consuming and slow to process, as well as the accuracy in determining the optimum result.

Keywords: quantum computing, quantum existence testing, finding extreme values in an unsorted database, resource management distribution, exhaustive algorithm, heuristic algorithm, computational complexity

1. INTRODUCTION

Nowadays, resource management and especially the distribution of resources is one of the most discussed and fundamental issues, for example, the problem considers a large amount of distributed energy resources including electric vehicles with gridable capacity [1], and the rapid progress of cloud computing, i.e. the growing number of video providers that have deployed their streaming services onto multiple distributed data centers [2]. This greatest demand on efficient resource distribution motivated us to search for new approaches and solutions.

The majority of operational processes require an amount of computational resources and their wide availability, namely the amount of computing units, is not the present issue, rather the utilization of resources and usage of the units are [3]. Thus, this challenge was tackled by introducing a new strategy which aims to optimize the computational load of the resources that handle the problems of computational complexity, time, speed and accuracy.

As is common knowledge, the primary aim of Quantum Computing and Communications is to reduce computational complexity and achieve optimum and efficient results with regard to the requirements of the given problem [4]. In order to exploit the power of quantum computing in terms of resource distribution, a quantum extreme value searching algorithm was used [5], which will be combined with an appropriately designed classical framework.

2. Modelling

Our novel model deals with the classical problem of resource distribution and contains three components.

The first one generates tasks to be computed in the system denoted by p_i and can be characterized as processing time, memory, energy, etc. The actual number of running tasks in the system is referred to as n. These tasks are served by computing units that represent the resources of the system. The number of the units is denoted by c. The computing units may have different theoretical capacities. The theoretical capacity of the j^{th} unit is depicted as s_j and its free (unused) capacity is denoted by $x_j[t]$ which depends on time t, as depicted in Fig. 1.

The third block is the decision-making unit which answers the question of how to deploy a new task among the process units in order to optimize the operation of the

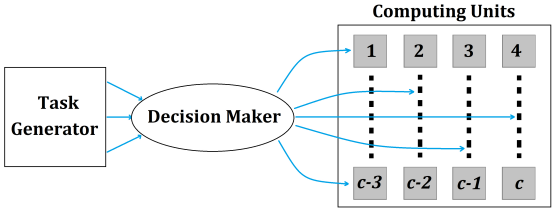


Figure 1: Architecture of the model

*Correspondence: elgaily@hit.bme.hu

system. There are different metrics to distribute the resources, here uniformly loaded units have been chosen because actually, they are an important aspect of many applications.

The overall capacity of the processing units in the system can be calculated as

$$\widehat{s} = \sum_{j=1}^{c} s_j, \tag{1}$$

while the overall free processing capacity is determined according to

$$\widehat{x} = \sum_{j=1}^{c} x_j[t]. \tag{2}$$

The average amount of free capacity per unit is given by the following expression

$$\overline{x}[t] = \frac{1}{c} \sum_{j=1}^{c} \frac{x_j[t]}{s_j}. \tag{3}$$

Our purpose is to uniformly distribute the load over the resources. Therefore, the variance of the relative free capacities in the system is used. In the case of optimal task distribution, if σ^2 tends to zero, then the resources are distributed uniformly, otherwise they are not. The corresponding formula of the relative variance is:

$$\sigma^2 = \frac{1}{c} \sum_{j=1}^{c} \left(\overline{x}[t] - \frac{x_j[t]}{s_j} \right)^2. \tag{4}$$

2.1 Description of the quantum algorithm

The optimized solution will be based on the quantum extreme value searching algorithm, a stochastic process which functions on an unsorted database that combines the technique of a classical binary search of a sorted database [6] and quantum existence testing (QET) [4]. The best classical solution requires N queries of the database to determine the optimum result, while in order to solve previous problems the well-known logarithmic (often referred to as binary) search algorithm, which is originally intended to search for a given item in a sorted database with quantum existence testing (a special case of quantum existence testing interested in whether a given entry exists in the database or not rather than in determining the number of existence entries) needed to be combined. The quantum algorithm maintains the efficiency of the binary search while processing an unsorted database [4].

It is hard to classically compute the optimum deployment scenario, therefore, the quantum extreme value searching algorithm [5] is applied as a minimum searching algorithm (MSA), which enables the deployment scenario to be identified and results in minimum variance. The MSA is a stochastic process that works on an unsorted database. Our new approach handles the database as a function, i.e. the variance. The proposed algorithm is now given in detail:

1: We start with $S = 0$:
$\sigma^2_{\min 1} = \sigma^2_{\min 0}$,
$\sigma^2_{\max 1} = \sigma^2_{\max 0}$,
and
$\Delta\sigma^2 = \sigma^2_{\max 0} - \sigma^2_{\min 0}$
2: $S = S + 1$
3: $\sigma^2_{\text{med } S} = \sigma^2_{\min S} + \left[\frac{\sigma^2_{\max S} - \sigma^2_{\min S}}{2} \right]$
4: $flag = \text{QET}\left(\sigma^2_{\text{med } S}\right)$
5: **if** $flag = \text{True}$ **then**
$\sigma^2_{\max S+1} = \sigma^2_{\text{med } S}$
6: **else**
$\sigma^2_{\max S+1} = \sigma^2_{\max S}$,
$\sigma^2_{\min S+1} = \sigma^2_{\text{med } S}$
7: **if** $S < \log_2 \left(\frac{\Delta\sigma^2}{\alpha} \right)$ **then**
goto 2
8: **else**
$y_{\text{opt}} = \sigma^2_{\text{med } S}$
stop

The algorithm will stop once the following step $S < \log_2 \left(\frac{\Delta\sigma^2}{\alpha} \right)$ has been fulfilled where α denotes the smallest sub-region between two possible results in a database and is explained in more detail in Section 2.3.

2.2 Description of the randomized, exhaustive and sequence-searching algorithms

The randomized, exhaustive and sequential- searching algorithms are generally viewed as references and the cornerstone with regard to finding solutions. They are considered as three different methods [7]. Firstly, the randomized approach anticipates the solution guided by given knowledge and is not perceived as an optimal solution because it randomly searches for one solution in one step from the space in which the duration of time may be less reasonable [8]. On the other hand, the exhaustive algorithm examines every possible solution and leads to an optimal result, it checks all the $\mathcal{O}(d)$ steps which is time-consuming and yields an accurate result. A large number of iterations are required when compared to the number of queries, hence, its time-consuming nature. Furthermore, the sequence method exhibits a similar degree of computational complexity to the randomized method, using $\mathcal{O}(\text{const})$ or $\mathcal{O}(1)$.

2.3 Evaluation of the algorithms

Our purpose is to provide a concrete comparison of solving distribution problems via heuristic 'random', exhaustive or sequence as well as quantum algorithms, hence, the difference in terms of the computational complexity applied and the minimum variance computed by each method was measured leading to the optimum distribution uniformity of the system.

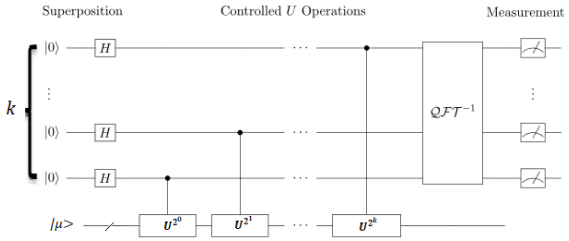


Figure 2: Quantum existing testing device (QFT: Quantum Fourier Transform).

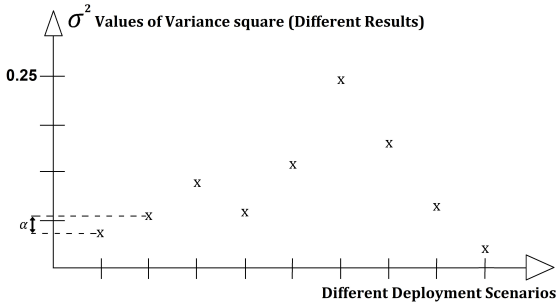


Figure 3: Functional representation of the database

Comparison according to computational complexity

By comparing the computational complexity of the quantum, randomized, exhaustive and sequence solutions, the quantum solution requires only $\sqrt{\log_2(T)\log_2^3\left(\sqrt{d}\right)}$ steps, where d refers to the number of different possible deployment scenarios and T denotes the number of sub-regions given by all the possible outcomes of the different deployments. The exhaustive solution, on the other hand, needs $\mathcal{O}(d)$. With regard to the randomized solution, it uses fewer steps so is faster than the three former solutions, which means that the computational complexity of the random method is less than for the other ones. Moreover, it does not determine the optimum solution since the computational complexity of the sequence method is similar to the randomized algorithm. In contrast, the exhaustive and quantum solutions require more computation steps, but the quantum solution always is preferred because it requires significantly less computation than the exhaustive method.

As was mentioned previously, the quantum extreme value searching algorithm uses the binary search and quantum existing testing methods. As the quantum phase estimation algorithm is the core of QET, as can be seen in Fig. 2, it outperforms the other counterpart's algorithms, thus, its physical implementation is highly constrained with the required number of bits k.

In fact, the number of bits k depends on the application of the system, as is illustrated in Fig. 2. This remains hard to realize, for example, if the error probability P_ε of the application is neglected and the classical specification considered, which is of accuracy a, the number of steps needed is $\mathcal{O}\left(\log_2(T)\log_2^3\left(\sqrt{d}\right)\right)$ and the number of bits will be influenced only by one factor, namely the accuracy a, as presented in

$$k = a - 1 \tag{5}$$

In addition, the relationship between the maximum relative variance and the number of sub-regions given T according to all the possible outcomes of the different deployments is given by

$$T = \frac{\sigma^2_{\max}}{\alpha} \tag{6}$$

where α denotes the smallest sub-region between two possible results in a database (Eq. 7). α is illustrated in Fig. 3:

$$\alpha = \min_{\forall\, i,j} \left|\left(\sigma_i^2 - \sigma_j^2\right)\right| \tag{7}$$

On the other hand, if two factors, namely the accuracy a and probability of error P_ε [4], are taken into consideration, this assumption will influence the second term of the computational complexity $\log_2^3\left(\sqrt{d}\right)$, it will be transformed into $\mathcal{O}\left(\log_2(T)\log_2^3\left(2^{\frac{P}{2}}\sqrt{d}\right)\right)$ and the number of bits will be expressed by

$$k = a - 1 + \underbrace{\left\lceil \log_2(2\pi) + \log_2\left(3 + \frac{1}{\check{P}_\varepsilon}\right)\right\rceil}_{P}, \tag{8}$$

where $\check{P}_\varepsilon$ is the maximally allowed quantum uncertainty (probability of error) and P is the number of qbit which controls the quantum uncertainty. The computational complexity of the classical (exhaustive) and quantum solutions is compared. The number of computational steps to yield the desired results with regard to the number of different deployment scenarios is in accordance with function $\mathcal{O}(d)$, while the quantum solution only requires $\mathcal{O}\left(\log_2(T)\log_2^3\left(\sqrt{d}\right)\right)$.

Furthermore, the quantum method derives its stochastic behavior from the quantum phase estimation algorithm which heightens the degree of accuracy and speed in computation.

In the light of what has been shown, in Fig. 4, the classical strategy for identifying the deployment scenario which leads to the minimum variance needs more computational computing, while the quantum solution requires significantly less computation, which reduces the complexity and duration necessary to determine the optimum deployment scenario of resource distribution.

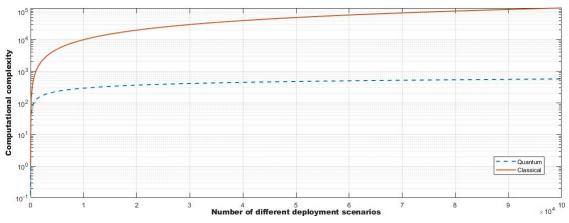


Figure 4: Comparison between the computational complexity of the classical and quantum decision-makers

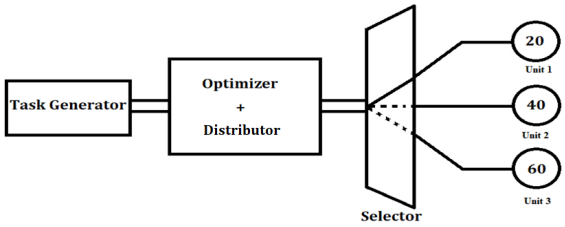


Figure 5: The simulation architecture

Comparison of the uniformity

In compliance with what has been discussed, the quantum and exhaustive solutions conserve the uniformity of the system, but the proposed quantum solution is the best and most efficient method because it requires less computation and time to determine the optimum deployment scenarios. However, the randomized and sequence algorithms do not ensure the uniformity of task distribution.

3. Results and Analysis

To show the importance of the proposed quantum solution, a simulation environment of Optimizer+Distributor was constructed.

The model of a resource distribution system contains three processing units with different theoretical capacities: twenty, forty and sixty running tasks in parallel. Practical systems contain significantly more processing units, however, to observe trends and effects it is worthwhile investigating a small-scale model first. Furthermore, a task generator block is present in the system. The tasks are considered to have exponential arrival times. An Optimizer+Distributor block is considered to be a decision-maker between the different deployment scenarios. The model implemented in the simulation environment is illustrated in Fig. 5.

3.1 Experiments

Two factors influence the behavior of the simulation: the mean (intensity) of the exponential distribution with regard to the arrival times of tasks, and the service time of the tasks in the processing units.

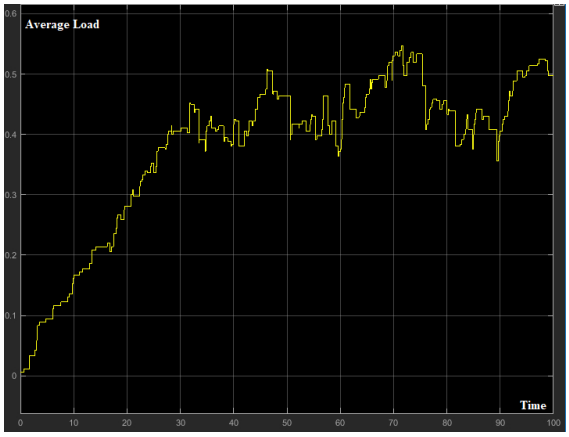


Figure 6: The average load of processing units in the case of the random reference strategy

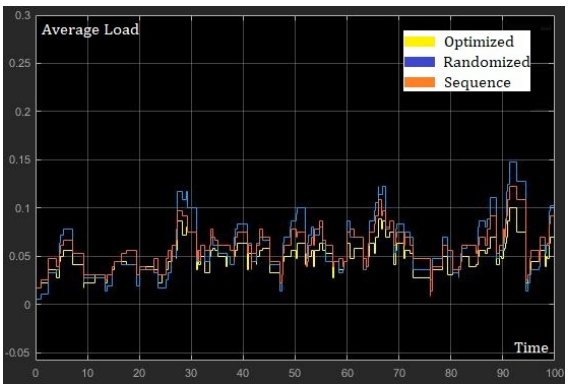


Figure 7: Average load of the three distribution strategies when the mean (intensity) of the exponential distribution of the arrival times of tasks is equal to 0.6

The system will be more heavily loaded if the mean value is smaller or the service time of the tasks larger.

3.2 Simulations

In order to demonstrate the efficiency of the proposed optimization strategy, two other reference strategies were considered which distribute the tasks randomly or sequentially among the processing units.

By considering the following simulation parameters, the mean of the exponential arrival times is equal to 0.4 and the service time of the resources is equal to 3.

The average load of the processing units in the case of the random reference strategy is presented in Fig. 6 with the following simulation parameters: the mean of the exponential arrival times is equal to 0.4 and the service time of the resources is equal to 15. The line graph contains two phases: the transition phase with a duration of 0 to 30 s, and the stationary phase that commences after approximately 30 s, when the system reaches a certain equilibrium.

Comparing the performance of the three distribution strategies, it can be stated that during the transition phase the variances of the reference systems are approximately stable, but during the stationary phase (normal operation) the variance started to fluctuate dramatically as well as increase. On the other hand, the variance of the proposed quantum solution remained approximately linear and tended to zero, therefore, the quantum system conserves the distribution uniformity.

According to Fig. 7 and 8 the average load of the different methods remains similar independent of the intensity of the exponential distribution of the arrival times of tasks as well as the decision methods. Furthermore, it is clearly noticeable that when the mean of the exponential distribution of the arrival times of tasks is smaller, the tasks are generated faster which leads to an increase in the load of the computing units.

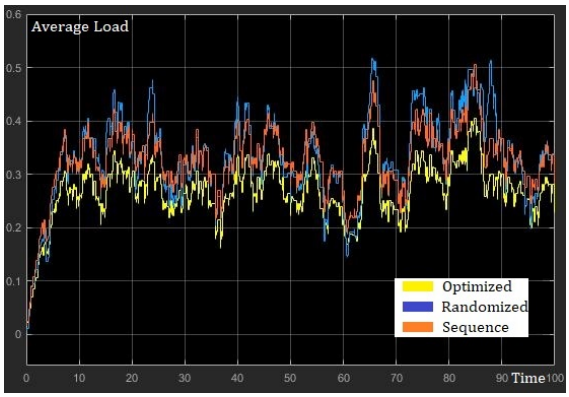


Figure 8: Average load of the three distribution strategies when the mean (intensity) of the exponential distribution of the arrival times of tasks is equal to 0.1

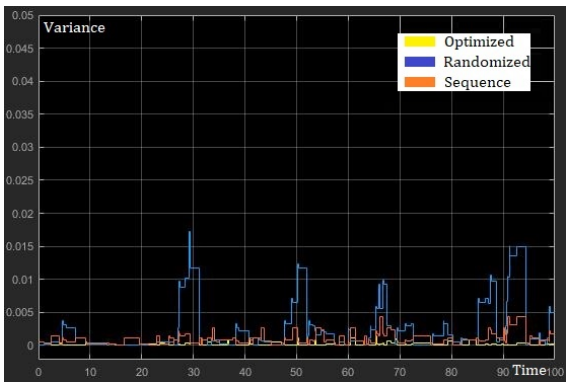


Figure 10: Variances over time of the optimized (yellow), randomized (blue) and sequence distribution (orange) strategies when the mean (intensity) of the exponential distribution of the arrival times of tasks is equal to 0.6

The simulation results concerning the variance of each distribution strategy show that the trends of the randomized and sequence strategies diverge from zero, but the sequence method yields a more uniform variance than the randomized one. In contrast, the optimized (quantum and exhaustive) strategies maintain and conserve the load uniformity of the system. These results are illustrated in Fig. 9 and 10.

4. Conclusion

In this paper, a new strategy for resource distribution based on a quantum searching algorithm was introduced. It was demonstrated that the quantum solution is more efficient by comparing the computational complexity and distribution uniformity of the quantum solution with the randomized, exhaustive and sequence methods. Furthermore, the changes applied to the computational complexity when the specification classical requirement depends on the application were demonstrated.

To show the importance of the quantum solution, a simulation environment of the proposed optimization of a distribution system was constructed and compared to two reference distribution systems which follow the randomized and sequence strategies. The proposed quantum approach is practical in most domains of infocommunication and computer science where resources have to be distributed among a large number of processing units.

These are the initial results of this new field of research. Obviously, other optimization metrics should be considered in the future. Furthermore, tasks can be modelled in a more sophisticated manner: different classes of tasks can be defined following various arrival processes and characterized by more than one resource parameter (processing time, memory, battery requirement). Finally, even an individual task can contain more blocks with dependencies among them and different blocks can be distributed among different processing units.

The wider future context of our research is to start with a polynomial time problem and progress towards a nondeterministic polynomial time problem.

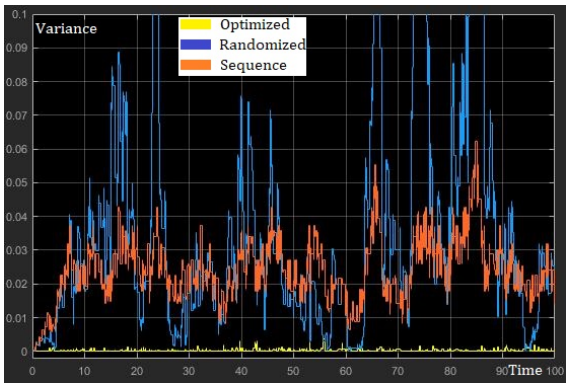


Figure 9: Variances over time of the optimized (yellow), randomized (blue) and sequence distribution (orange) strategies when the mean (intensity) of the exponential distribution of the arrival times of tasks is equal to 0.1

Acknowledgement

The research was partially supported by the National Research Development and Innovation Office of Hungary (Project No.2017-1.2.1-NKP-2017-00001), by the Hungarian Scientific Research Fund -OTKA K-112125 and in part by BME Artificial Intelligence FIKP grant of EMMI (BME FIKP- MI/SC). The research reported in this paper was supported by the BME- Artificial Intelligence FIKP grant of EMMI (BME FIKP-MI/SC).

REFERENCES

[1] Soares, J.; Canizes, B.; Vale, Z; Venayagamoorthy, G. K.: Benders' decomposition applied to energy resource management in smart distribution networks, Clemson University Power Systems Conference, Clemson, SC. March 8, 2016. DOI: 10.1109/PSC.2016.7462820

[2] Zhang, Z.; Jiang, X.; Xi, H.: Joint resource allocation and traffic management for cloud video distribution over software-defined networks, 8th IEEE International Conference on Communication Software and Networks, University of Science and Technology of China, Hefei, China. 2016. DOI: 10.1109/ICCSN.2016.7586692

[3] Keränen, J.; Kortelainen, J.; Antila, M.: Computational resource management system. Version 1.0/14.8.2015. https://www.vtt.fi/inf/julkaisut/muut/2015/VTT-R-01975-15.pdf

[4] Imre, S.; Balázs, F.: Quantum computing and communications: An engineering approach (Wiley, Chichester, UK) 2005 ISBN 0-470-86902-X, DOI: 10.1002/9780470869048

[5] Imre, S.: Extreme value searching in unsorted databases based on quantum computing, *Int. J. Quantum Inf.,* 2005 **3**(1), 171–176 DOI: 10.1142/S0219749905000700

[6] Knuth, D. E.: Sorting and searching in The art of computer programming Vol. 3 (2nd ed.) (Addison-Wesley, Reading, MA, USA) 1998 ISBN 978-0-201-89685-5

[7] Zin, N. A. M.; Abdullah, S. N. H. S; Zainal, N. F. A.; Ismail, E.: A comparison of exhaustive, heuristic and genetic algorithm for travelling salesman problem in PROLOG, *Int. J. Adv. Sci. Eng. Inf. Technol.,* 2012 **2**(6), 459–463 DOI: 10.18517/ijaseit.2.6.244

[8] Hooker, J. N.: Toward unification of exact and heuristic optimization methods, *Int. T. Oper. Res.,* 2013 **22**, 19–48 DOI: 10.1111/itor.12020

HUNGARIAN JOURNAL OF
INDUSTRY AND CHEMISTRY
Vol. 47(1) pp. 17–23 (2019)
hjic.mk.uni-pannon.hu
DOI: 10.33927/hjic-2019-04

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

SIMULATION OF COLOR AFTERIMAGES: AN APPROACH TO COMPUTING VIRTUAL COLOR PERCEPTION

LŐRINC GARAI *[1] AND ANDRÁS HORVÁTH[2]

[1]Doctoral School of Multidisciplinary Engineering Sciences (MMTDI), Széchenyi István University, Egyetem tér 1, Győr, 9026, HUNGARY
[2]Department of Physics and Chemistry, Széchenyi István University, Egyetem tér 1., Győr, 9026, HUNGARY

Afterimages are a common and frequent perceptual phenomenon of everyday life. When looking into a high-intensity light source and suddenly turning away from it, a temporary "ghost" of the light source remains visible, for a while. The computer-graphics simulation of afterimages is based on biophysical and mathematical models as published in the literature. A subordinate of afterimages defined in our research is *virtual color perception,* that is in our interpretation an unusual and intense temporary color perception provoked by a rapid change in the color of the incident light. In research, the modelling of virtual color perception is a field that is by and large untouched. Our publication presents a kinetic model established to characterize the intensity and duration of virtual color perception as a function of rapid changes in the color of the incident light.

Keywords: afterimage, rod and cone photoreceptors, photopigment, kinetics

1. INTRODUCTION

In our vision, an afterimage is an illusionary image that appears after having been exposed to a prior one. Color afterimages are experienced in everyday life, for example, when driving at night the headlights of oncoming cars are so bright that when the driver looks away from them, the illusion of bright headlights still remains in perception [1]. When photorealistic images are rendered [1–4], some papers reported simulations of color afterimages by combining mathematical models [5–9].

A subordinate of afterimages defined in our research is *virtual color perception*, a phenomenon that originates from chromatic adaptation in photoreceptors influenced by environmental color interactions, and based on the sensitivity to light and adaptibility of each cone receptor.

The physiological background of virtual color perception in brief is as follows: human photopic (daylight) color vision is a combined response of type L (long wavelengths), M (medium wavelengths) and S (short wavelengths) cone receptors to adequate stimuli of light. The photopigment *rhodopsin* plays a key role in the process; the equilibrium of its relative concentration is achieved by the opposing processes of rapid cleavage when exposed to light and slow resynthesis in darkness [10, 11]. Adaptations of cone receptors to changes in the color of the incident light is time-consuming [12, 13], hence, a rapid change in the color of the incident light facilitates unusual and intense temporary color perception, the so-called *virtual color perception*. For example, when exposed to red light the sensitivity of L cone receptors is low and in this case is accompanied by the high sensitivity of M and S cone receptors. Following a rapid change in color from red to blue in the incident light, the sensitivity of S cone receptors remains temporarily high, resulting in perception of the color bright blue that transforms into common blue after a short period of time during which the relative concentration of photopigment is equilibrated (restored), i.e. this is the duration of *virtual color perception*.

The purpose of our work is to develop a computational kinetic model capable of simulating and quantifying virtual color perception.

In our research chromaticity diagrams are used. Note that changes in the xy coordinates [14] are not proportional to human color perception. To overcome this distortion, several chromaticity coordinates were defined, for example, CIE $u'v'$ [15].

The gamut of a device is the complete subset of colors it can produce. Usually, in an RGB (red, green, blue) device, it consists of a color triangle with two-dimensional chromaticity coordinates. A gamut is characteristic of the given display (screen) currently in use, for example, modern RGB LED displays are characterized by wider gamuts compared to old-fashioned cold cathode fluores-

*Correspondence: garailorinc@garailorinc.hu

Table 1: Gamut points

	CCFL		RGB LED	
Gamut point	x	y	x	y
B	0.2091	0.2218	0.1563	0.0307
B3R1	0.2753	0.2518	0.2837	0.1056
B2R2	0.3415	0.2817	0.4111	0.2181
B1R3	0.4077	0.3115	0.5385	0.3024
R	0.4739	0.3415	0.6658	0.3305
R3G1	0.4420	0.3837	0.5687	0.4236
R2G2	0.4101	0.4239	0.4717	0.5632
R1G3	0.3783	0.4652	0.3746	0.6679
G	0.3464	0.5064	0.2775	0.7028
G3B1	0.3121	0.4353	0.2472	0.5348
G2B2	0.2778	0.3641	0.2169	0.2827
G1B3	0.2434	0.2930	0.1866	0.0937

cent lamp (CCFL) displays that were frequently used about 10 years ago.

In our work, gamuts characteristic of CCFL and RGB LED desktop monitors were measured first. Following this, the simulation of virtual color perception obtained by gamut data as a result of a rapid change in the color of the incident light was conducted. Finally, preliminary validation tests were run on the aforementioned RGB LED desktop monitor in use [16].

2. Experimental

2.1 *Measurement of the gamuts of the displays used in our experiments and key parameters of our model*

The spectral power distribution of the red, green and blue primaries of two displays was measured by a spectroradiometer (a Flame Miniature Spectrometer by Ocean Optics, Inc. calibrated 12 strong lines of He, Ne, Ar and H_2 flashtubes). One of the displays used was that of an old notebook using a CCFL as a backlight and the other was a more modern one (HP ZR2440w) with a display using RGB LEDs as a backlight. Based on the spectral power distributions measured, the CIE 1931 (x, y) chromaticity coordinates were calculated for all three primaries of both displays, using a Color Matching Function (CMF) of 10° at a resolution of 1 nm between the wavelengths of 360 nm and 830 nm. Intermediate gamut point coordinates were calculated by interpolation. Gamut point numbers in Table 1 and Fig. 1 were further referred to as colors of incident light. In our kinetic model, actual color perception is compiled from the generally known mathematical relations [8, 9, 17] shown below.

The actual color perception J of a single (L, M or S) cone receptor can be calculated by the formula

$$J = DEp, \tag{1}$$

where D denotes a conversion constant between the cleavage of rhodopsin and neural impulses and here is

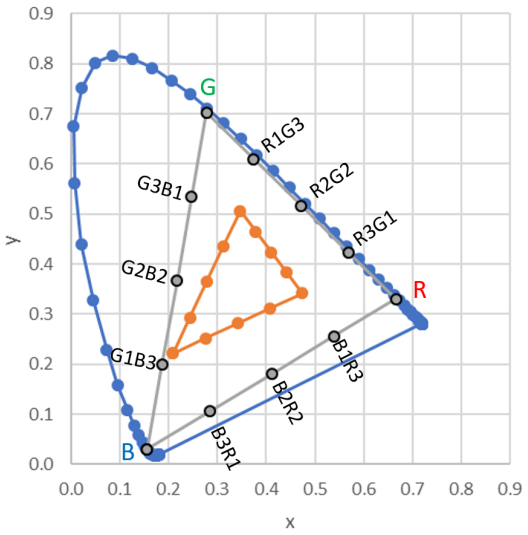


Figure 1: Chromaticity diagram of gamut points: an old CCFL display of a notebook (inner gamut) and an HP ZR2440w display (outer gamut)

equal to 1. The variable E represents the intensity of incident light expressed in trolands (Td). During the calculations a maximum luminance of the monitor of 300 cd/m^2 was used and a diameter of the pupil of 5 mm assumed. Therefore, the maximum retinal illuminance was equal to 5890 Td. The variable p denotes the relative concentration of photopigment (between 0 and 1).

The time differential of p determined from the rate of photopigment synthesis (Q_s), spontaneous photopigment cleavage (Q_c) and photoinduced cleavage (Q_i) is calculated by

$$\frac{\mathrm{d}p}{\mathrm{d}t} = Q_\mathrm{s} - Q_\mathrm{c} - Q_\mathrm{i}. \tag{2}$$

The variables of Eq. 2 are calculated by

$$Q_\mathrm{s} = \frac{1}{\tau}, \tag{3}$$

$$Q_\mathrm{c} = \frac{p}{\tau}, \tag{4}$$

and

$$Q_\mathrm{i} = \frac{E}{E_0}\frac{p}{\tau}, \tag{5}$$

where the time constant $\tau = 99$ 1/s and $E_0 = 20,000$ are used [13]. The following differential equation is composed from Eqs. 2–5:

$$\frac{\mathrm{d}p}{\mathrm{d}t} = \frac{1}{\tau} - \frac{p}{\tau} - \frac{E}{E_0}\frac{p}{\tau}. \tag{6}$$

The solution of Eq. 6 yields the actual relative concentration of photopigment:

$$p(t) = \frac{1}{b}\left[1 - (1 - p_0 b)\right] e^{-tb/\tau}, \tag{7}$$

where p_0 denotes the initial relative concentration of photopigment.

Finally, the equilibrium with regard to the relative concentration of photopigment p_e and percentage of photopigment cleaved b are related as follows:

$$p_e = \frac{1}{b} = \frac{E_0}{E + E_0}. \tag{8}$$

2.2 Simulation formula

In accordance with CIE 1931 [18, 19], the actual coordinates of color perception $x(t)$, $y(t)$ and $z(t)$ are calculated from the color coordinates of incident light x_i, y_i and z_i by equations Eqs. 9–24. In the equations below, variables indexed with L, M and S apply to cone receptors L, M and S, respectively.

$$E_L = E_{multip} \cdot M_{1,1\text{-}3} \times [x_i, y_i, z_i] \tag{9}$$

$$E_M = E_{multip} \cdot M_{2,1\text{-}3} \times [x_i, y_i, z_i] \tag{10}$$

$$E_S = E_{multip} \cdot M_{3,1\text{-}3} \times [x_i, y_i, z_i] \tag{11}$$

$$b_L = b_M = b_S = 1 + \frac{E_{multip}}{E_0} \tag{12}$$

E_{multip}, which is equal to $6,000$, denotes the light intensity of the display. The actual relative concentration of photopigment is calculated from the initial relative concentrations of photopigment p_{0L}, p_{0M} and p_{0S}:

$$p_L(t) = \frac{1}{b_L}(1 - (1 - p_{0L}b))e^{-tb/\tau} \tag{13}$$

$$p_M(t) = \frac{1}{b_M}(1 - (1 - p_{0M}b))e^{-tb/\tau} \tag{14}$$

$$p_S(t) = \frac{1}{b_S}(1 - (1 - p_{0S}b))e^{-tb/\tau} \tag{15}$$

For the purposes of iteration in simulations, the initial relative concentration of photopigment p_0 was equal to 0.1. From the actual relative concentrations of photopigment Eqs. 13–15, the color perception coordinates are as follows

$$J_L(t) = D \cdot p_L \cdot E_L, \tag{16}$$

$$J_M(t) = D \cdot p_M \cdot E_M, \tag{17}$$

$$J_S(t) = D \cdot p_S \cdot E_S, \tag{18}$$

where J_L, J_M and J_S denote the cone receptors of long, medium and short wavelengths, respectively.

To obtain more accurate color perception coordinates, tristimulus values were calculated:

$$X(t) = M^{-1}_{1,1\text{-}3} \times [J_L, J_M, J_S], \tag{19}$$

$$Y(t) = M^{-1}_{2,1\text{-}3} \times [J_L, J_M, J_S], \tag{20}$$

$$Z(t) = M^{-1}_{3,1\text{-}3} \times [J_L, J_M, J_S], \tag{21}$$

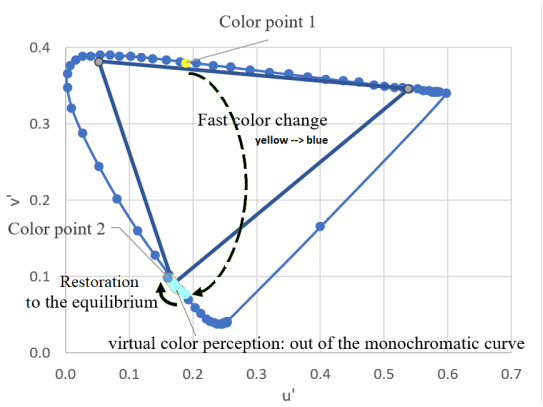


Figure 2: An example of fast color change leading to virtual color perception. Color point yellow shows primary color perception. Color points bright blue show virtual color perception as reflected by the tendency to reach equilibrium in photopigment relative concentration.

where M denotes a transformation matrix between tristimulus values X, Y and Z, and the actual color perception J. The actual color perception coordinates $x(t)$, $y(t)$, $z(t)$ are calculated by the following equations:

$$x(t) = \frac{X(t)}{X(t) + Y(t) + Z(t)}, \tag{22}$$

$$y(t) = \frac{Y(t)}{X(t) + Y(t) + Z(t)}, \tag{23}$$

$$z(t) = 1 - x(t) - y(t). \tag{24}$$

In accordance with the CIELUV (1976) chromaticity diagram, the actual color perception coordinates $x(t)$, $y(t)$ and $z(t)$ are transformed into coordinates $u'(t)$ and $v'(t)$ by an easy-to-compute method [19] :

$$u'(t) = \frac{4x(t)}{12y(t) - 2x(t) + 3}, \tag{25}$$

$$v'(t) = \frac{6y(t)}{12y(t) - 2x(t) + 3}. \tag{26}$$

The *intensity* of the actual virtual color perception is determined by

$$\Delta c = \sqrt{(u'(t) - u'_e)^2 + (v'(t) - v'_e)^2}, \tag{27}$$

where u'_e and v'_e denote color coordinates at equilibrium following restoration from virtual color perception (see *Restoration to the equilibrium* in Fig. 2).

A summary of variables and parameters is shown in Notations at the end of this paper.

2.3 Simulation

To understand the calculations, a graphical approach is shown in Fig. 2. Gamut color point 1 stands for the primary perception of the actual incident light, which is yellow here. With a rapid change in color from yellow to

Table 2: Example of iteration

$t(s)$	x_i	y_i	$u'(t)$	$v'(t)$	Δc
0.0	0.4250	0.56875	–	–	–
10.0	0.4250	0.56875	0.1893	0.3802	–
20.0	0.4250	0.56875	0.1892	0.3802	–
30.0	0.4250	0.56875	0.1891	0.3802	–
30.1	0.1400	0.05000	0.1748	0.0836	0.02075
30.2	0.1400	0.05000	0.1748	0.0837	0.02066
30.3	0.1400	0.05000	0.1747	0.0838	0.02057
30.4	0.1400	0.05000	0.1747	0.0838	0.02048
30.5	0.1400	0.05000	0.1746	0.0839	0.02039
30.6	0.1400	0.05000	0.1745	0.0840	0.02031
30.7	0.1400	0.05000	0.1745	0.0840	0.02022
30.8	0.1400	0.05000	0.1744	0.0841	0.02013
30.9	0.1400	0.05000	0.1744	0.0842	0.02004
31.0	0.1400	0.05000	0.1743	0.0842	0.01995
31.1	0.1400	0.05000	0.1743	0.0843	0.01987
31.2	0.1400	0.05000	0.1742	0.0844	0.01978
…	…	…	…	…	…
63.9	0.1400	0.05000	0.1611	0.0991	0.000126
64.0	0.1400	0.05000	0.1611	0.0991	0.000101
64.1	0.1400	0.05000	0.1611	0.0992	0.000092
64.2	0.1400	0.05000	0.1611	0.0992	0.000102

blue, a bright blue color appears in perception that transforms into common blue after a short period of time necessary for the restoration of the equilibrium in terms of the relative concentration of photopigment, which is the time period required for virtual color perception, namely for the perception of bright blue (Fig. 2).

Our kinetic simulation model (Section 2.2) is illustrated in Table 2. The first five lines in the first four columns show the same values of the color coordinates of incident light $x_\mathrm{i}, y_\mathrm{i}, z_\mathrm{i}$, against time ($0 - 30$ seconds).

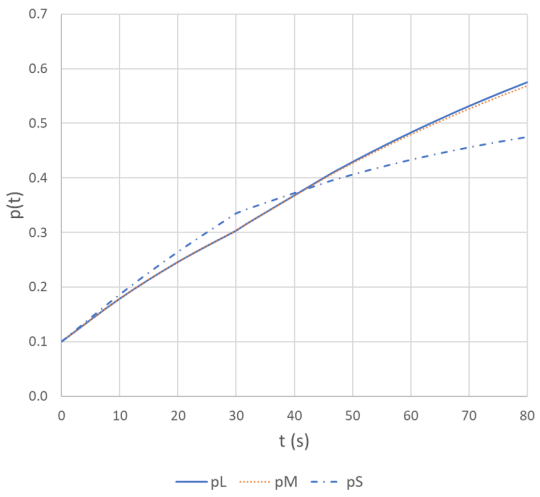


Figure 3: Photopigment relative concentration values in the iteration in Table 2

Over this 30 second-long period, the $u'(t)$ and $v'(t)$ values of color perception are quasi identical. However, after 30 seconds, a rapid change in color of the incident light from red to blue results in virtual color perception, as demonstrated by line 6 and column 5. The values of Δc in column 7 concern the intensity of virtual color perception. $(\Delta c)_\mathrm{max}$ denotes the peak intensity in virtual color perception and the minimum Δc stands for the duration of virtual color perception (t_virtcol). Actual relative concentrations of photopigment of cone receptors L, M and S are shown in the diagram in Fig. 3.

In terms of simulating virtual color perception, rapid changes in the color of incident light were indicated on CCFL and RGB LED monitors by the assignment of defined gamut points to each other (Fig. 1). Altogether, 24 changes in color were simulated.

2.4 Validation of the simulation

Our kinetic model was validated by a test that consisted of 20 subjects involving an in-house piece of software run in a Python environment. Accordingly, a homogeneous solid colored circle is displayed on a homogeneous background of a different color for 30 seconds (Fig. 4), then the circle disappears (Fig. 5) and the intensity and duration of virtual color perception is determined by the key inputs of the user. Further details concerning the test are found below:

- First, the test subject looked at a white screen for 30 seconds.

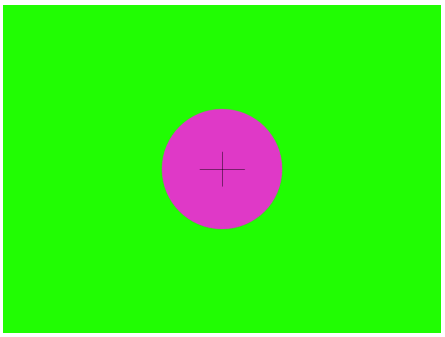


Figure 4: Second screen of validation test.

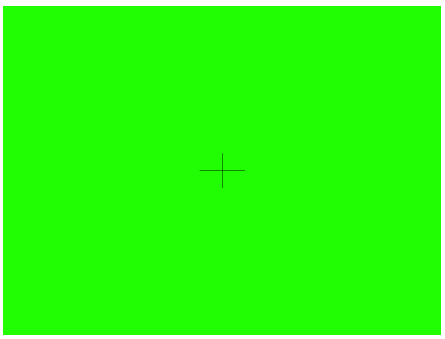


Figure 5: Third screen of validation test: circle removed.

Table 3: Comparison of virtual color perception intensity in model situations and in validation test results

Ranking in model		Ranking category	Ranking in validation test results		
ID	Rank		ID	Median of ranking	Correlation
R → G	1	High	B → G	1.00	No
G2B2 → R	2		B2R2 → G	2.00	Yes
G1B3 → R	3		R → G	3.00	Yes
B2R2 → G	4		B → R2G2	4.00	No
R → G2B2	5	Medium	R → G2B2	4.00	Yes
R2G2 → B	6	Low	G1B3 → R	4.50	No
G → B2R2	7		G2B2 → R	5.00	No
B → G	8		G → B2R2	5.00	Yes
B → R2G2	9		R2G2 → B	7.00	Yes

Matching percentage between model and validation test results: 56%

Table 4: Comparison of virtual color perception time period in model situations and in validation test results

Ranking in model		Ranking category	Ranking in validation test results		
ID	Rank		ID	Median of ranking	Correlation
R → G2B2	1	High	B → G	1.00	No
R → G	2		G1B3 → R	2.00	No
G → B2R2	3		B → R2G2	3.00	No
B2R2 → G	4		R → G2B2	4.00	Yes
R2G2 → B	5	Medium	G2B2 → R	4.00	No
G2B2 → R	6	Low	G → B2R2	4.50	No
B → R2G2	7		B2R2 → G	5.00	No
G1B3 → R	8		R → G	5.00	No
B → G	9		R2G2 → B	7.00	No

Matching percentage between model and validation test results: 11%

- Second, the eyes of the subject were fixed on a circle at the center of the subsequent colored image (Fig. 4) for 30 seconds. According to our definition, the color of the central circle represents the first gamut point, while the color of the background represents the second gamut point. Altogether, 9 assignments of gamut points have been validated so far.
- Third, the central circle suddenly disappeared (Fig. 5) and the subject responded according to the intensity and duration of virtual color perception experienced.
- The intensity of virtual color perception was rated on a four-grade scale, where zero stands for the absence of virtual color perception and 4 denotes its highest intensity. The duration of virtual color perception was indicated by the subject pressing a key as the perception faded away.

First, the assignments of gamut points for each test subject were ranked according to the intensity (Table 3) and duration (Table 4) of virtual color perception induced. Then, the median of the rank order with regard to the intensity and duration of virtual color perception was calculated.

Table 5: Intensity and time period of virtual color perception related to gamut points assignments

		$(\Delta c)_{\max}$		$t_{\mathrm{virtcol}}(s)$	
Color point 1	Color point 2	CCFL	RGB LED	CCFL	RGB LED
B	R3G1	0.00792	0.01212	30.9	24.5
B	R2G2	0.00553	0.00513	36.3	27.5
B	R1G3	0.00339	0.00197	41.9	16.0
R3G1	B	0.00794	0.01210	33.5	29.5
R2G2	B	0.00852	0.01367	34.9	32.2
R1G3	B	0.00953	0.01523	37.2	34.6
G3B1	R	0.00931	0.02381	27.1	29.3
G2B2	R	0.00958	0.02187	27.5	27.6
G1B3	R	0.00994	0.01989	28.1	25.8
R	G3B1	0.00920	0.02298	46.9	44.2
R	G2B2	0.00821	0.01686	51.2	52.8
R	G1B3	0.00735	0.01260	50.2	38.2
G	B3R1	0.00807	0.01789	48.2	41.1
G	B2R2	0.00648	0.01362	45.8	42.2
G	B1R3	0.00693	0.01722	30.5	31.0
B3R1	G	0.00324	0.01337	25.0	27.8
B2R2	G	0.00524	0.01799	33.7	33.7
B1R3	G	0.00749	0.02256	41.4	38.7
B	R	0.01036	0.01786	28.8	23.8
R	B	0.00788	0.01053	33.3	26.6
R	G	0.00980	0.02707	48.2	43.1
G	R	0.00915	0.02571	26.7	30.8
G	B	0.01084	0.01677	40.1	37.0
B	G	0.00234	0.00868	17.8	20.5

Since the number of test subjects was limited, the results could not be divided according to their age and gender. Further tests are needed to ensue virtual color perception with regard to gender and age.

3. Results and Discussion

In Table 5, the maximum $(\Delta c)_{\max}$ for the CCFL and RGB LED displays were identified during the rapid change in the color of incident light from gamut point G to B (from green to blue) and R to G (from red to green), respectively.

When compared to the CCFL display, the RGB LED display appears to yield higher $(\Delta c)_{\max}$ values with the exception of rapid changes in the color of the incident light from gamut point B to R2G2 (from blue to orange).

With regard to our results, the duration of virtual color perception seems to be platform-free, i.e. t_{virtcol} displayed on both the CCFL and RGB LED monitors was identical. However, rapid changes in the color of the incident light from gamut point B to R1G3 (from blue to yellowish green) was an exception with regard to the values of t_{virtcol}. As is shown in Table 5, t_{virtcol} computed on the CCFL display has doubled in value compared to that computed on the RGB LED display.

The kinetic simulation results so far point to the likelihood of the appearance of virtual color perception on all

display platforms.

As for our preliminary tests performed on 20 test subjects so far as well as the parameters $(\Delta c)_{\max}$ and t_{virtcol}, a correlation between model situations (simulations) and the results of a validation test cannot be confirmed at present. Further tests, statistical evaluations and the introduction of additional parameters are also necessary to achieve more accurate conclusions.

4. Conclusion

Photopic human color vision is a combined response to the stimulation from light of red, green and blue cone receptors. Cone receptors adapt individually to the actual color of the incident light. Since the adaptation of cone receptors is time-consuming, virtual color perception can be achieved in the meantime by rapid changes in the color of the incident light.

Our kinetic model developed for individual cone receptors is based on mathematical correlations that simulate the intensity and duration of virtual color perception which result from rapid changes in color. According to our model, virtual color perception can result from both CCFL and RGB displays.

Our preliminary validations are yet to confirm a correlation between model situations (simulations) and the results of a validation test. Some refinements to the simulation by the introduction of additional parameters as well as further validation tests with regard to the gender and age of participants are indispensable to reach more accurate conclusions.

Acknowledgement

This research would not have been possible without the participation of volunteers from Széchenyi István University as test subjects.

Notations

Symbol	Meaning	Unit
J	in actual color perception: J_{L} = red cone receptors specific for long wavelength J_{M} = green cone receptors specific for medium wavelength J_{S} = blue cone receptors specific for short wavelength	dimensionless
p	relative photopigment concentration	dimensionless (range:$0 \ldots 1$)
D	conversion constant between rhodopsin cleavage and neural impulses	dimensionless (value: 1)
E	incident light intensity	troland
Q_{s}	Synthesis of photopigment	dimensionless
Q_{c}	Spontaneous cleavage of photopigment	dimensionless
Q_{i}	Light induced cleavage of photopigment	dimensionless
τ	time constant in rhodopsin synthesis	seconds
p_0	photopigment initial relative concentration	dimensionless
$p(t)$	photopigment relative concentration at t time after p_0	dimensionless
b	percentage of photopigment cleaved	dimensionless
M	transformation matrix of tristimulus values XYZ and the actual color perception J	dimensionless

REFERENCES

[1] Mikamo, M.; Slomp, M.; Raytchev, B.; Tamaki, T.; Kaneda, K.: Perceptually inspired afterimage synthesis, *Computers & Graphics,* 2013 **37**(4), 247–255 DOI: 10.1016/j.cag.2013.02.008

[2] Gutierrez, D.; Anson, O.; Munoz, A.; Seron, F. J.: Perception-based rendering: eyes wide bleached in Dingliana, J.; Ganovelli, F. (eds): Eurographics 2005 - Short Presentations (The Eurographics Association, Geneve, Switzerland) 2005 pp. 49–52 DOI: 10.2312/egs.20051021

[3] Ritschel, T.; Eisemann, E.: A computational model of afterimages, *Comput. Graph. Forum,* 2012 **31**(2pt3), 529–534 DOI: 10.1111/j.1467-8659.2012.03053.x

[4] Horváth, A; Dömötör, G.: Computational simulation of mesopic vision based on camera recordings, *Light and Engineering,* 2014 **22**(1), 61–67

[5] Reidenbach, H.-D.: Determination of the time dependence of colored afterimages in Stuck, B. E.; Belkin, M.; Manns, F.; Söderberg, P. G.; Ho, A. (eds): Proceedings Ophthalmic Technologies XVIII **6844**, 2008 DOI: 10.1117/12.762852

[6] Padgham, C. A.: Quantitative study of visual afterimages, *Brit. J. Ophthal.,* 1953 **37**, 165–170 DOI: 10.1136/bjo.37.3.165

[7] Padgham, C. A.: Measurements of the colour sequences in positive visual after-images, *Vision Res.,* 1968 **8**(7), 939–949 DOI: 10.1016/0042-6989(68)90142-9

[8] Alpern, M.: Rhodopsin kinetics in the human eye, *J. Physiol.,* 1971 **217**(2), 447–471 DOI: 10.1113/jphysiol.1971.sp009580

[9] Smith, C. V.; Pokorny, J.; Van Norren D.: Densitometric measurement of human cone photopigment kinetics, *Vision Res.,* 1983 **23**(5), 517–524 DOI: 10.1016/0042-6989(83)90126-8

[10] Linksz, A.: An essay on color vision and clinical color-vision tests, *Am. J. Ophthalmol.,* 1964 **58**(3), 513 DOI: 10.1016/0002-9394(64)91250-4

[11] Hurvich, L. M.: Color Vision, (Sinauer Associates Inc., Sunderland, USA), 1981 ISBN: 978-0878933365

[12] Ábrahám, Gy.; Kovács, G.; Antal, Á.; Németh, Z.; Veres, Á. L.: Jármű optika (in Hungarian), 2014 http://www.mogi.bme.hu/TAMOP/jamu_optika/ch02.html#ch-II.3.3.2.2 (downloaded on: 02/10/2018)

[13] Szentágothai J.: Functionalis anatomia, (in Hungarian) (Medicina Kiadó, Budapest, Hungary), 1977, vol. 3, pp. 1590–1597

[14] Colorimetry, 3rd Edition, CIE 15:2018 ISBN: 978 3 901906 33 6

[15] http://www.color-theory-phenomena.nl/10.03.htm (downloaded on: 02/10/2018)

[16] Csuti P.: The characterization of the photometry and colorimetry of light emitting diodes, PhD dissertation (in Hungarian) DOI: 10.18136/PE.2016.642 (http://konyvtar.uni-pannon.hu/doktori/2016/Csuti_Peter_dissertation.pdf)

[17] Horváth A.: A fényterjedés és -észlelés fizikája mérnököknek (in Hungarian), (SZE-MTK, Fizika és Kémia Tanszék, Győr) 2013, pp. 165-178, ISBN: 978-963-7175-97-8

[18] Smith T.; Guild J.: The C.I.E. colorimetric standards and their use, *Trans. Opt. Soc.*, 1932 **33**(3), 73–134 DOI: 10.1088/1475-4878/33/3/301

[19] Schanda J.: Colorimetry: understanding the CIE system (John Wiley & Sons, Inc., New Jersey, USA), 2007 DOI: 10.1002/9780470175637

HUNGARIAN JOURNAL OF
INDUSTRY AND CHEMISTRY
Vol. 47(1) pp. 25–32 (2019)
hjic.mk.uni-pannon.hu
DOI: 10.33927/hjic-2019-05

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

THERMAL MODEL DEVELOPMENT FOR A CUBESAT

NAWAR AL HEMEARY *[1,2], MACIEJ JAWORSKI[3], JAN KINDRACKI[3], AND GÁBOR SZEDERKÉNYI[1]

[1]Faculty of Information Technology and Bionics, Pázmány Péter Catholic University, Práter u. 50/A, Budapest, 1083, HUNGARY
[2]Electromechanical Engineering Department, University of Technology, Al Senaha St., Baghdad, 10066, IRAQ
[3]Faculty of Power and Aeronautical Engineering, Warsaw University of Technology, Nowowiejska 24, Warszawa, 00-665, POLAND

CubeSats provide a cost-effective means of several functions of satellites due to their small size, mass, relative simplicity and short development time. Therefore, CubeSat technologies have been widely studied and developed by space organizations, companies and educational institutions all over the world. These satellites have certain drawbacks. Small surface areas are a consequence of their small size which often imply thermal and power constraints. A novel development of CubeSats known as PW-Sat has been developed by Warsaw University of Technology. A control-oriented lumped thermal model of this satellite containing a fuel tank in the form of nonlinear ordinary differential equations is proposed in this paper. The model is able to simulate the thermal behavior of the surface and fuel tank of the satellite in its orbit. For the PW-Sat to operate reliably, the temperature of the fuel tank has to be maintained within given safety limits. Because of the limited power, passive thermal control is assumed in this case. Several simulation results are presented for different surface compositions to determine whether they are able to guarantee the prescribed temperature range throughout the entire orbit or not.

Keywords: PW-Sat, TMM, thermal behavior, propellant tank, dynamical modeling

1. INTRODUCTION

In recent years, interest in Cube-Satellites (CubeSats) has grown tremendously within the space community from space agencies as well as in industry and academia. Two factors have influenced this spurt of interest, namely the low-cost nature of access to space and the utilization of commercial off-the-shelf (COTS) technologies in the design architecture. These two factors have led to a significantly low overall cost of a CubeSat mission [1]. A CubeSat is cubic in form with edges 10 cm in length and a mass of up to 1.33 kg. The PW-Sat is a CubeSat that has been in development for more than a year by different teams at Warsaw University of Technology [2, 3]. At present, PW-Sat has never been flown with an onboard propulsion system. Due to the significant and growing interest in CubeSat mission capabilities, several propulsion systems have been rapidly developed for use in CubeSats such as cold gas propulsion systems, solar sails, electric propulsion systems and chemical propulsion systems [4, 5]. Cold gas propulsion systems are relatively simple solutions in CubeSats. Gas from a high-pressure gas cylinder is simply vented through a valve and nozzle to produce thrust [6, 7].

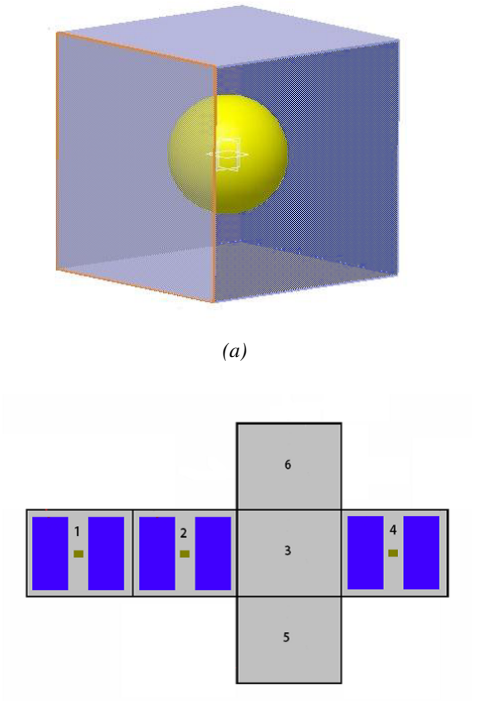
*Correspondence: al.hemeary@itk.ppke.hu

The goal of this paper is to propose a lumped dynamical model of temperatures during orbital motion in the main parts of a CubeSat that contains a fuel tank. The model is built in the form of nonlinear ordinary differential equations (ODEs). Although advanced thermal simulation tools exist that utilize detailed distributed mathematical models, this simple form of a model has been chosen since it is intended to be used in the model of temperature control design. The overwhelming majority of control design techniques require models in the form of ordinary differential equations [8]. Moreover, in the case of nonlinear models such as the CubeSat system studied, a low dimensional model is preferred due to the computational complexity of control design. This approach is also supported by control theory and practice, namely that in general such simple models are sufficient for controller design [9]. As a first step in terms of the regulation of temperature, passive control in the form of the appropriate composition of materials covering the surfaces of the satellite is used. During the construction of the model, the standard principles of thermal modeling [10, 11] and their application in aerospace engineering are followed [12, 13].

Relevant results can be found in the literature concerning the thermal modeling and analysis of small satel-

lites in the form of ODEs. In [14] a simple thermal dynamical model of a CubeSat containing two differential equations is presented. The two lumped balance volumes are the surface and internal parts of the satellite, respectively. It is shown that the problem is mathematically analogous to the forced vibration of a damped mechanical system. In [9], new theoretical results on the qualitative behavior of spacecraft thermal models are provided that contain several nodes (compartments). It is proven that such models exhibit a unique asymptotically stable equilibrium in the positive orthant with constant external disturbance inputs which leads to a stable limit cycle during orbital motion. The analysis concerning the frequency domain of a multi-compartmental model of a satellite is conducted in [15]. The ODEs are linearized around the equilibrium points which permit the use of Fourier analysis.

The structure of the paper is as follows: the description starts with the derivation of a simple mathematical model to simulate the transient thermal behavior of the fuel tank as well as the satellite faces in orbit. Then, by changing either the optical properties of the surface of the PW-Sat or the solar cell ratio of the satellite surface, several simulations are presented to illustrate how different configurations satisfy the given temperature limits.



(a)

(b)

Figure 1: PW-Sat Structure ((a) PW-Sat spherical fuel tank set, (b) PW-Sat solar cell arrangements).

2. SYSTEM DESCRIPTION AND ASSUMPTIONS

The intended use of the model developed in this paper is twofold: 1) to study the effect of different surface compositions (including solar cells) on temperature, 2) to evaluate the possibility of installing a propellant tank in the CubeSat. With regard to the modeling, the following assumptions are made:

2.1 Structural Assumption

PW-Sat is a cubic structural bus with a total mass of 1 kg composed of six faces as walls. The basic structure of these faces is composed of the aluminum alloy 6061-T6 with various optical surface properties. These properties are based on uncoated surfaces for one experiment and coated with a magnesium oxide-aluminum oxide paint for the others. The nitrogen fuel tank, made of stainless steel with a diameter of 5 cm, is planned to be installed in the center of this satellite as shown in Fig. 1a and is assumed to contain an internal gas subject to 100 bars of pressure at an initial temperature of 298 K.

The solar cells will be attached to sides 1, 2 and 4 of the PW-Sat, as shown schematically in Fig. 1b, because these faces will be exposed to solar radiation due to the assumed orbit of this satellite.

2.2 Orbital Assumption

The PW-Sat is designed for a circular low Earth orbit (LEO). The total orbital period (P) is 1.5 h. However, the motion of PW-Sat is assumed to be identical when exposed to solar radiation and during shadow passage at an altitude of 300 km and an inclination of zero. Face 3 is directed towards the Earth throughout its orbit. Faces 1, 2 and 4 are exposed to the sun with regard to the orbital motion of the satellite. Finally, faces 5 and 6 are directed towards the space along the satellite orbit as shown in Fig. 2.

2.3 Thermal Assumption

The six faces of the PW-Sat are considered to have a uniform temperature distribution. The conductive heat transfer between the fuel tank and satellite is ignored to simplify the thermal modeling calculations. Only face 3 is exposed to infrared radiation from the Earth in this orbit and the albedo during the luminous orbit intervals

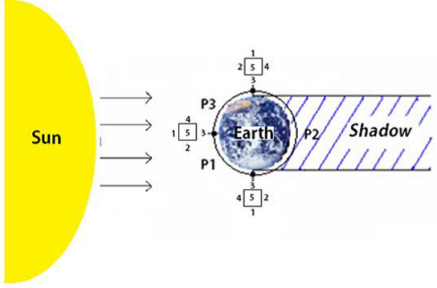


Figure 2: PW-Sat orbital motion.

[14, 16]. The thermal rate of power dissipation generated from the operation of elements from the satellite is assumed to be 2 W. The thermal limits of the fuel tank are 228 K and 338 K, and for the surface area of the satellite are 173 K and 373 K.

3. THERMAL MATHEMATICAL MODEL (TMM)

The problem concerns the formulation of the model, which is on the one hand simple enough to limit the expenditure, on the other hand, detailed enough to present an adequate description of the physical surroundings [17]. The main purpose of these calculations is to divide the periodic motion of the satellite (with period P) into three intervals (parts) as shown in Fig. 2. The first interval (P_1) starts with an initial time of $t = 0$ s when face 1 faces the sun and ends at a time of $t = 1350$ s when face 4 faces the sun. The second interval (P_2) is an eclipse interval between $t = 1351$ s and $t = 4050$ s. The third interval (P_3) starts at $t = 4051$ s when face 2 faces the sun and ends at the end of the satellite period at $t = 5399$ s.

3.1 First Interval Equations

$$\text{Interval } P_1 : t = 0 \rightarrow t = \frac{P}{4} \tag{1}$$

The satellite spends a quarter of its orbital period in this luminous part. During this time interval, the surface of the satellite receives direct solar and albedo radiation depending on the position of the satellite due to its motion and the satellite emits thermal IR radiation into space; however, only face 3 receives additional infrared radiation from Earth because it faces the Earth [18]. The rate of heat transfer between face 1, the external environment and the spherical fuel tank during the first interval can be described by

$$\begin{gathered}(m_{\mathrm{Al}}\, C_p + m_{\mathrm{sc}}\, C_p^{\mathrm{sc}})\frac{dT_1}{dt} = \\ G_{\mathrm{s}} a_{\mathrm{s}}^{\text{Al-sc}} A \cos\left(\frac{2\pi t}{P}\right) + \dot{Q} + \dot{Q}_{F_1} - \varepsilon_{\mathrm{IR}}^{\text{Al-sc}} \sigma A T_1^4\end{gathered} \tag{2}$$

where m_{Al} denotes the mass of aluminum, C_p stands for the specific heat of aluminum (980 J/(kg·K)), m_{sc} represents the mass of the solar cell , C_p^{sc} is the specific heat of the solar cell (1600 J/(kg·K)), T_1 denotes the temperature of face 1, G_{s} stands for the solar constant (1367 W/m^2), and $a_{\mathrm{s}}^{\text{Al-sc}}$ represents the average solar absorptance of aluminum and the solar cell which is calculated as (Al %·$a_{\mathrm{s}}^{\mathrm{Al}}$+sc %·$a_{\mathrm{s}}^{\mathrm{sc}}$), where Al % and sc % denote the percentages of aluminum and solar cells in the cover, respectively. A stands for the surface area of the face (0.01 m^2), $\dot{Q}$ represents the thermal rate of power dissipation, $\dot{Q}_{F_1}$ denotes the radiative heat transfer between face 1 and the tank, $\varepsilon_{\mathrm{IR}}^{\text{Al-sc}}$ is the average infrared emissivity of Al and sc which is calculated as (Al %·$\varepsilon_{\mathrm{Al}}$+sc %·$\varepsilon_{\mathrm{sc}}$) and σ stands for the Stefan-Boltzmann constant ($5.669 \cdot 10^{-8}$W/m^2 K^4) [6].

The rate of heat transfer between face 2, the external environment and the spherical fuel tank during the first interval can be described by

$$\begin{gathered}(m_{\mathrm{Al}}\, C_p + m_{\mathrm{sc}}\, C_p^{\mathrm{sc}})\frac{dT_2}{dt} = \\ \dot{Q} + \dot{Q}_{F_2} - \varepsilon_{IR}^{\text{Al-sc}} \sigma A T_2^4\end{gathered} \tag{3}$$

where T_2 denotes the temperature of face 2 and $\dot{Q}_{F_2}$ stands for the radiative heat transfer between face 2 and the tank.

The rate of heat transfer between face 3, the external environment and spherical fuel tank during the first interval can be described by

$$\begin{gathered}m\, C_p \frac{dT_3}{dt} = AF \cdot G_{\mathrm{s}} a_{\mathrm{s}}^{\mathrm{Al}} F_{\mathrm{sE}} A \cos\left(\frac{2\pi t}{P}\right) + \\ \dot{Q} + \dot{Q}_{F_3} + a_{\mathrm{IR}}^{\mathrm{Al}} \sigma A T_{\mathrm{E}}^4 - \varepsilon_{\mathrm{IR}}^{\mathrm{Al}} \sigma A T_3^4\end{gathered} \tag{4}$$

where m denotes the mass of the face (0.04 kg), F_{sE} stands for the view factor between the face of the satellite and the Earth which is almost one [18], T_3 represents the temperature of face 3, $\dot{Q}_{F_3}$ is the radiative heat transfer between face 3 and the tank, AF denotes the factor on the albedo (0.28), $a_{\mathrm{s}}^{\mathrm{Al}}$ stands for the solar absorptivity of aluminum, $a_{\mathrm{IR}}^{\mathrm{Al}}$ represents the infrared absorptivity of aluminum, T_{E} is the reference temperature of the Earth (255 K) and $\varepsilon_{\mathrm{IR}}^{\mathrm{Al}}$ denotes the infrared emissivity of aluminum.

The rate of heat transfer between face 4, the external environment and the spherical fuel tank during the first interval can be modeled as

$$\begin{gathered}\left(m_{\mathrm{Al}}\, C_p + m_{\mathrm{sc}}\, C_p^{\mathrm{sc}}\right)\frac{dT_4}{dt} = \\ G_{\mathrm{s}} a_{\mathrm{s}}^{\text{Al-sc}} A \sin\left(\frac{2\pi t}{P}\right) + \dot{Q} + \dot{Q}_{F_4} - \varepsilon_{\mathrm{IR}}^{\mathrm{Al}} \sigma A T_4^4\end{gathered} \tag{5}$$

where T_4 denotes the temperature of face 4 and $\dot{Q}_{F_4}$ stands for the radiative heat transfer between face 4 and the tank.

The rate of heat transfer between face 5, the external environment and the spherical fuel tank during the first interval can be described as

$$m\, C_p \frac{dT_5}{dt} = \dot{Q} + \dot{Q}_{F_5} - \varepsilon_{\mathrm{IR}}^{\mathrm{Al}} \sigma A T_5^4 \tag{6}$$

where T_5 denotes the temperature of face 4 and $\dot{Q}_{F_5}$ stands for the radiative heat transfer between face 5 and the tank.

The rate of heat transfer between face 6, the external environment and the spherical fuel tank during the first interval can be described as

$$m\, C_p \frac{dT_6}{dt} = \dot{Q} + \dot{Q}_{F_6} - \varepsilon_{\mathrm{IR}}^{\mathrm{Al}} \sigma A T_6^4 \tag{7}$$

where T_6 denotes the temperature of face 6 and $\dot{Q}_{F_6}$ stands for the radiative heat transfer between face 6 and the tank.

3.2 Second Interval Equations

$$\text{Interval } P_2 : t = \frac{P}{4} \rightarrow t = \frac{3}{4}P \tag{8}$$

These concern the duration of an eclipse. The satellite spends half of its orbital period in an eclipse. During this interval, the surface of the satellite receives neither direct solar nor albedo radiation, whilst face 3 still receives IR radiation from the Earth because it faces it. The satellite emits thermal IR radiation into space. Therefore, the rates of heat transfer of the faces of the satellite (1, 3 and 4) have slightly changed in their equations compared to the first interval.

The rate of heat transfer between faces 1, 3 and 4 as well as the external environment during the second interval can be described by the following equations

$$(m_{\mathrm{Al}}\, C_p + m_{\mathrm{sc}}\, C_p^{\mathrm{sc}})\frac{dT_1}{dt} = \dot{Q} + \dot{Q}_{F_1} - \varepsilon_{\mathrm{IR}}^{\text{Al-sc}}\sigma A T_1^4 \tag{9}$$

$$m\, C_p \frac{dT_3}{dt} = \dot{Q} + \dot{Q}_{F_3} + a_{\mathrm{IR}}^{\mathrm{Al}}\sigma A T_{\mathrm{E}}^4 - \varepsilon_{\mathrm{IR}}^{\mathrm{Al}}\sigma A T_3^4 \tag{10}$$

$$\begin{aligned}(m_{\mathrm{Al}}\, C_p + m_{\mathrm{sc}}\, C_p^{\mathrm{sc}})\frac{dT_4}{dt} = \\ \dot{Q} + \dot{Q}_{F_4} - \varepsilon_{\mathrm{IR}}^{\text{Al-sc}}\sigma A T_4^4\end{aligned} \tag{11}$$

3.3 Third Interval Equations

$$\text{Interval } P_3 : t = \frac{3}{4}P \rightarrow t = P \tag{12}$$

The satellite spends a quarter of its orbital period in this second luminous part. During this time interval, the surface of the satellite receives and emits thermal radiation in a similar manner to during interval 1 with only a slight change in the equation of face 2.

The rate of heat transfer between this face, the external environment and the spherical fuel tank during the third interval can be modeled as

$$\begin{aligned}(m_{\mathrm{Al}}\, C_p + m_{\mathrm{sc}}\, C_p^{\mathrm{sc}})\frac{dT_2}{dt} = \\ G_{\mathrm{s}} a_{\mathrm{s}}^{\text{Al-sc}} A \left|\sin\left(\frac{2\pi t}{P}\right)\right| + \\ \dot{Q} + \dot{Q}_{F_2} - \varepsilon_{IR}^{Al-sc}\sigma A T_2^4\end{aligned} \tag{13}$$

3.4 The transient heat transfer of the spherical propellant tank

The rate of heat transfer between the faces of the satellite and the fuel tank can be described by the following equation

$$(m_{\mathrm{s}}\, C_p^{\mathrm{s}} + m_{\mathrm{g}} C_V)\frac{dT_{\mathrm{t}}}{dt} = -\sum_{1}^{6} \dot{Q}_{F_n} \tag{14}$$

Table 1: Material properties of PW-Sat.

	Al 6061-T6 uncoated	Al 6061-T6 coated	Solar cells
Specific Heat [J/(kg·K)]	980	980	1600
Emissivity (thermal)	0.08	0.92	0.85
Absorptivity (solar)	0.379	0.09	0.92

Table 2: Material properties of fuel tank.

	Stainless Steel	Nitrogen
Specific Heat [J/(kg·K)]	504	743
Mass [kg]	0.0926	0.0074

Table 3: Average of optical surface properties of partially covered surfaces.

Cube Face	ε	a	Coverage
1,2 and 4	0.89	0.33	70 % Al, 30 % sc
	0.87	0.67	30 % Al, 70 % sc
3,5 and 6	0.92	0.09	Al painted

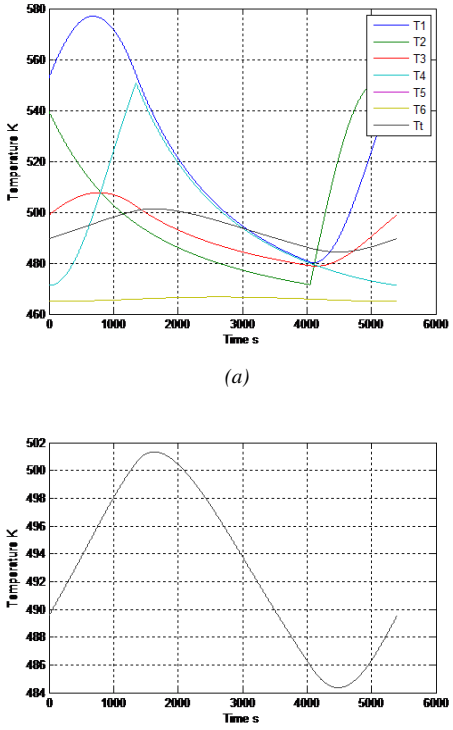
where m_{s} denotes the mass of stainless steel, C_p^{s} stands for the specific heat of stainless steel (504 J/(kg·K)), m_{g} represents the mass and C_V the specific heat of nitrogen (743 J/(kg·K)), and T_{t} is the temperature of the tank.

The radiative heat transfer between each face and the tank $\dot{Q}_{F_n}$, depending on which face it applies to, can be described as

$$\dot{Q}_{F_n} = F_{\mathrm{ft}}\varepsilon\sigma A(T_{\mathrm{t}}^4 - T_n^4) \tag{15}$$

The view factor between the face in question and the fuel tank is given by $F_{\mathrm{ft}} = \frac{1}{(1+H)^2}$ (see, e.g. [10, 11]), where H denotes the ratio of the distance between the spherical surface of the tank to the surface of the internal face ($h = 0.025$ m) in terms of the radius ($r = 0.025$ m), which is expressed as $H = \frac{h}{r}$.

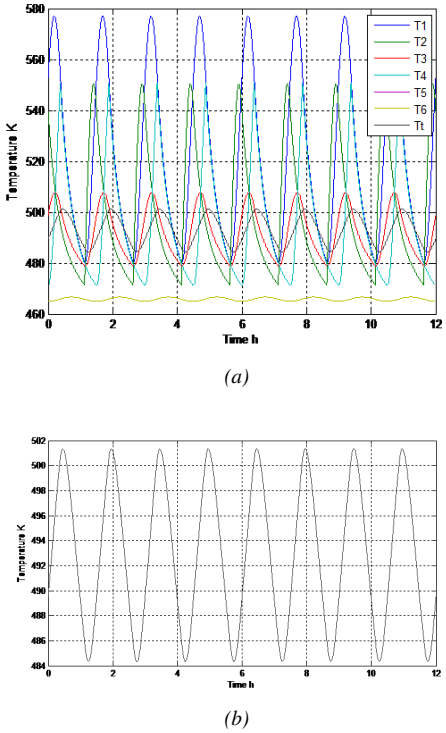
The mass of the solar cell m_{sc} per unit area is on average 850 g/m^2, thus, the mass of the solar cell as a proportion of the total mass of the face is determined by the equation ($m_{\mathrm{sc}} = 850$ g/m$^2 \cdot A \cdot$ sc %). The total mass of the tank m_{t} is assumed to be 0.1 kg, so the mass of nitrogen gas was calculated by assuming the initial temperature and total pressure of the tank. The optical surface properties are shown in Table 1, and the masses of both nitrogen gas and the tank are shown in Table 2. Hence, it is necessary to calculate the properties of the average materials to conduct a thermal analysis. The emissivity of infrared radiation from these faces can be calculated as the average of the emissivity of infrared radiation from aluminum and the emissivity of infrared radiation from the solar cell in addition to the absorptivity of these faces as shown in Table 3.



(a)

(b)

Figure 3: The thermal behavior using 100 % Al during one orbital period ((a) temperatures of the faces and fuel tank, (b) temperature of the fuel tank) (T1, . . . , T6) refers to the temperatures of faces 1, . . ., 6, respectively, and (T_t) denotes the temperature of the tank.



(a)

(b)

Figure 4: Thermal behaviors using 100 % Al during 8 orbital periods ((a) temperatures of the faces and fuel tank, (b) temperature of the fuel tank).

4. COMPUTATIONAL RESULTS

In this section, the faces and investigations into the thermal behavior of the tank are presented for several cases based on: uncoated surfaces, surfaces coated with magnesium oxide-aluminum oxide paint and different feasible options of the ratios of solar cells from the PW-Sat to simulate the temperature of the fuel tank with different optical properties of the surface materials and solar cell ratios.

4.1 The PW-Sat faces composed of 100 % uncoated aluminum

In this case, the thermal simulations of the faces and fuel tank were conducted according to the assumption that the faces of the satellite are composed of the aluminum alloy 6061-T6. Starting with the equations of the intervals and by using the ODE45 solver in MATLAB (the simulation time step was 1 s), the thermal behavior during the orbital motion of the satellite was computed.

1 - The thermal simulation of the faces and spherical fuel tank during one orbital period (time span from 0 s to 5399 s) is shown in Fig. 3a and the simulation of the temperature of the tank during this orbital period is shown in Fig. 3b. It can be seen that the predefined temperature limits are not adhered to in this case, since the minimum temperatures of the faces of the satellite and fuel tank exceed 460 K during its orbital period.

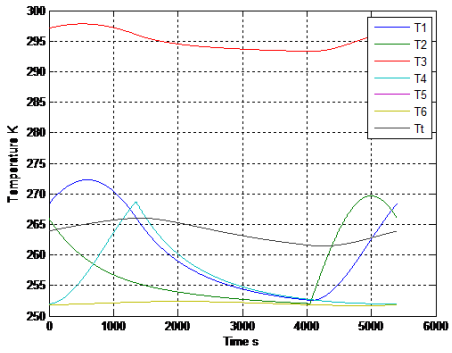
2 - The thermal simulation of the faces and spherical fuel tank during several orbits (8 orbital periods with a time span of 12 h to illustrate long-term operations) is shown in Fig. 4a, and the simulation of the temperature of the tank during these orbital periods is shown in Fig. 4b.

4.2 The faces of the PW-Sat are composed of 100 %-coated aluminum

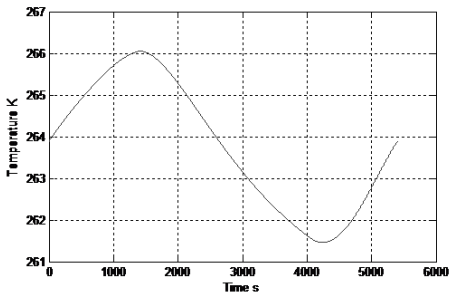
The thermal simulations of the faces and fuel tank were conducted according to the assumption that the surface of the satellite was composed of aluminum coated with magnesium oxide-aluminum oxide paint. By using the assumed time span of each interval, the results are shown below:

1 - The thermal simulations of the faces and spherical fuel tank during one orbital period (time span from 0 s to 5399 s) are shown in Fig. 5a, and the simulation of the temperature of the tank during this orbital period is shown in Fig. 5b. It can be seen that all the defined temperature limits are adhered to in this case.

2 - The thermal simulation of the faces and spherical fuel tank over 8 orbital periods (time span of 12 h) is shown in Fig. 6a, and the simulation of the temperature of the tank during these orbital periods is shown in Fig. 6b.

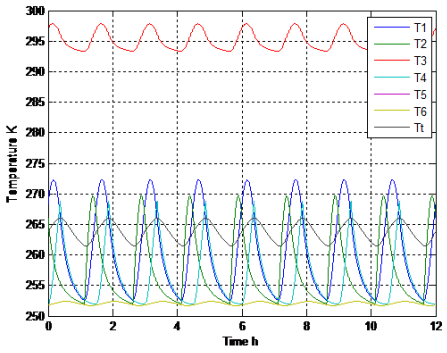


(a)

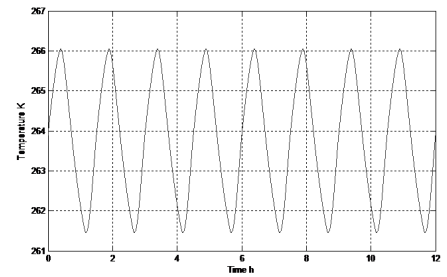


(b)

Figure 5: Thermal behaviors using coated Al during one orbital period ((a) temperatures of the faces and fuel tank, (b) temperature of the fuel tank).



(a)



(b)

Figure 6: Thermal behaviors using coated Al over 8 orbital periods ((a) temperatures of the faces and fuel tank, (b) temperature of the fuel tank).

4.3 *The faces of the satellite exposed to the sun during its orbit are covered with 70 % aluminum and 30 % solar cells*

In this case, these three sides of the PW-Sat are assumed to be composed of 70 % aluminum and 30 % solar cells and the other faces are coated with magnesium oxide-aluminum oxide paint. The simulation results are as follows:

1 - The thermal simulation of the faces and spherical fuel tank during one orbital period (time span of 0 s to 5399 s) is shown in Fig. 7a and the simulation of the temperature of the tank during this orbital period is also shown separately in Fig. 7b. The results show that the temperature of the tank varied between a maximum of 281.9 K and a minimum of 265.4 K, while the temperatures of the faces also remained within the given temperature limits.

2 - The thermal simulation of the faces and spherical fuel tank over 8 orbital periods (time span of 12 h) is shown in Fig. 8a, and the simulation of the temperature of the tank during these orbital periods is shown in Fig. 8b.

4.4 *The faces of the satellite exposed to the sun during its orbit are covered with 30 % aluminum and 70 % solar cells*

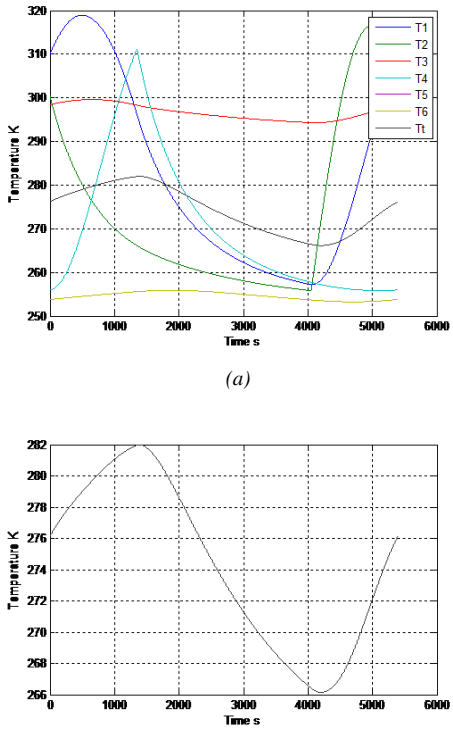
The results, according to the assumption that three sides of the PW-Sat are composed of 30 % aluminum and 70 % solar cells while the rest of them are coated with magnesium oxide-aluminum oxide paint, are shown below:

1 - The thermal simulation of the faces and spherical fuel tank during one orbital period (time span from 0 s to 5399 s) is shown in Fig. 9a, while simulation of the temperature of the tank during this orbital period is shown in Fig. 9b. The results show that the temperature of the tank varied between a maximum of 302.4 K and a minimum of 270.6 K, while the temperatures of the faces were within thermal limits.

2 - The thermal simulation of the faces and spherical fuel tank over 8 orbital periods (time span of 12 h) is shown in Fig. 10a, while the simulation of the temperature of the tank during these orbital periods is shown in Fig. 10b.

5. Conclusion

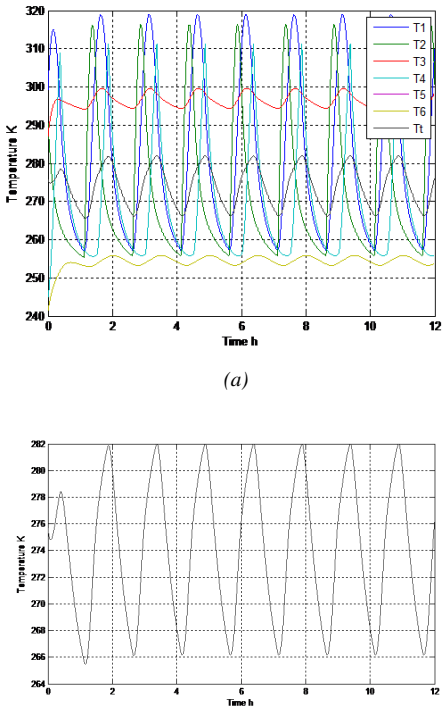
A thermal mathematical model was constructed and studied to compute the temperatures of the surfaces and fuel tank of the PW-Sat. Several cases were presented using various surface compositions and the results show that the proposed TMM is able to calculate the radiative heat



(a)

(b)

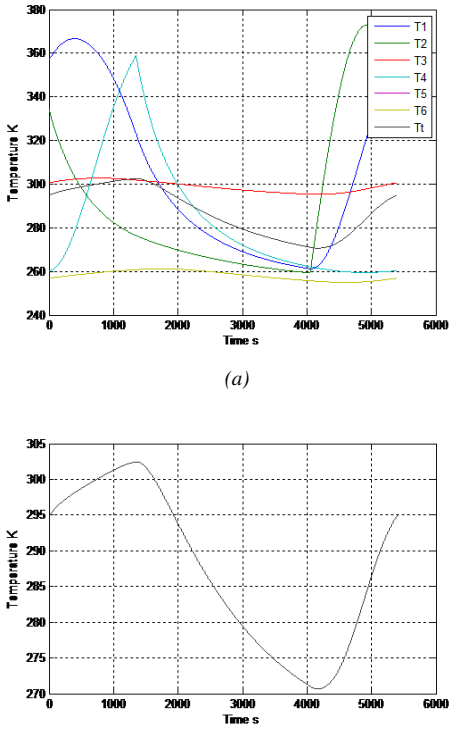
Figure 7: Thermal behaviors with 30 % sc during one orbital period ((a) temperatures of the faces and fuel tank, (b) temperature of the fuel tank).



(a)

(b)

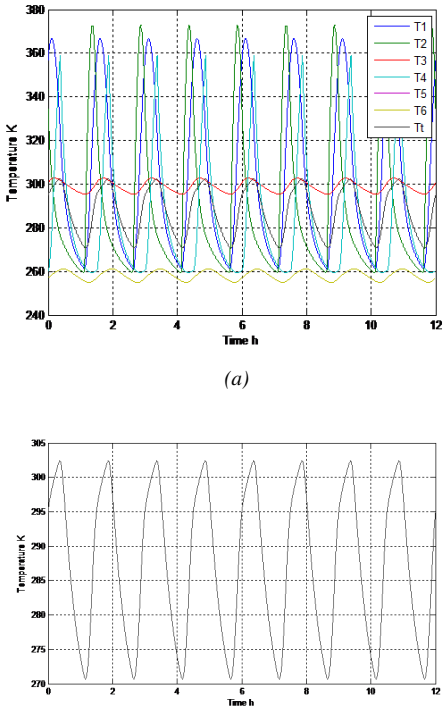
Figure 8: Thermal behaviors with 30 % sc over 8 orbital periods ((a) temperatures of the faces and fuel tank, (b) temperature of the fuel tank).



(a)

(b)

Figure 9: Thermal behaviors with 70 % sc during one orbital period ((a) temperatures of the faces and fuel tank,(b) temperature of the fuel tank).



(a)

(b)

Figure 10: Thermal behaviors with 70 % sc over 8 orbital periods ((a) temperatures of the faces and fuel tank, (b) temperature of the fuel tank).

that the PW-Sat would encounter during its assumed orbit. Initially, the surfaces of the PW-Sat were assumed to be composed of 100 % of the aluminum alloy 6061-T6. The corresponding results suggest that the temperatures of the surfaces and fuel tank would be too high. Therefore, additional finishes applied to the surfaces were taken into consideration. The first choice of finish was to coat the entire surface of the satellite with magnesium oxide-aluminum oxide paint. The obtained results show that the temperatures of the surfaces and fuel tank dropped because of the increasing emissivity and decreasing absorptivity of the surfaces. Further simulations were performed of cases in which the faces were exposed to the sun when partially covered with solar cells. The results indicate that the case described in Subsection 4.4, which delivers the most electrical power due to the highest percentage of solar cells, still satisfies the temperature limits of the fuel tank and surfaces of the satellite. The results also suggest that it is possible to install a fuel tank inside the PW-Sat which could be the first step required to add a propulsion system that can generate thrust for this CubeSat. Future works will include validation of the model using advanced thermal simulation tools as well as active control design to precisely regulate the temperature of the fuel tank.

Acknowledgement

This work has been partially supported by the European Union, co-financed by the European Social Fund through the grant EFOP-3.6.3-VEKOP-16-2017-00002. The support of the University of Technology of Baghdad and Warsaw University of Technology is also acknowledged.

REFERENCES

[1] Tummala, A. R.; Dutta, A.: An overview of cube-satellite propulsion technology and trends, *Aerospace,* 2017 **4**(4), 58. DOI: 10.3390/aerospace4040058

[2] Mehrparvar, A.: CubeSat design specification Rev. 13, The CubeSat Program, Cal Poly SLO, 2015. http://www.cubesat.org/resources

[3] PW-Sat description (CubeSat Warsaw University of Technology). https://directory.eoportal.org/web/eoportal/satellite-missions/p/pw-sat

[4] Lemmer, K.: Propulsion for CubeSats, *Acta Astronaut.,* 2017 **134**, 231–243. DOI: 10.1016/j.actaastro.2017.01.048

[5] Gaite, J.: Nonlinear analysis of spacecraft thermal models, *Nonlinear Dyn.* 2011 **65**, 283–300. DOI: 10.1007/s11071-010-9890-4

[6] Sutton, G. P.; Biblarz, O.: Rocket propulsion elements, Seventh Edition (Wiley, New York, USA) 2001. ISBN 0-471-32642-9

[7] Mueller, J.; Hofer, R.; Parker, M.; Ziemer, J.: Survey of propulsion options for CubeSats, 57th JANNAF Propulsion Meeting (Colorado) 2010 JANNAF-1425 1–56. https://trs.jpl.nasa.gov/bitstream/handle/2014/41627/10-1646.pdf

[8] Isidori, A.: Nonlinear Control Systems (Springer, London, UK) 1999. ISBN 978-1-84628-615-5

[9] Hangos, K. M.; Bokor, J.; Szederkényi, G.: Analysis and Control of Nonlinear Process Systems (Springer, London, UK) 2004. ISBN 978-1-85233-861-9

[10] Modest, M. F.: Radiative Heat Transfer, Third Edition (Academic Press, Oxford, UK) 2013. ISBN 9780123869906

[11] Sala, A.: Radiation Heat Transfer (in polish) (WNT, Scientific and Technical Publications, Warsaw, Poland) 1982. ISBN 8320403677

[12] Gilmore, D. G.: Spacecraft thermal control handbook, Second Edition, vol. 1. (The Aerospace Press, El Segundo, USA) 2002. ISBN 1-884989-11-X

[13] Diaz-Aguado, M. F.; Greenbaum, J.; Fowler, W. T.; Lightsey, E. G.: Small satellite thermal design, test, and analysis, Modeling, Simulation, and Verification of Space-based Systems III, *Proc. SPIE,* 2006 **6221**, 622109. DOI: 10.1117/12.666177

[14] Grande, I. P.; Andres, A. S.; Guerra, C.; Alonso, G.: Analytical study of the thermal behaviour and stability of a small satellite, *Appl. Therm. Eng.,* 2009 **29**(11-12), 2567–2573. DOI: 10.1016/j.applthermaleng.2008.12.038

[15] Farrahi, A.; Pérez-Grande, I.: Simplified analysis of the thermal behavior of a spinning satellite flying over Sun-synchronous orbits, *Appl. Therm. Eng.* 2017 **125**, 1146–1156. DOI: 10.1016/j.applthermaleng.2017.07.033

[16] JAXA: Design standard, Spacecraft thermal control system, JERG-2-310 (Japan Aerospace Exploration Agency, Tsukuba, Japan) 2009. http://sma.jaxa.jp/en/TechDoc/Docs/E_JAXA-JERG-2-310_08_RE.pdf

[17] Fortoscue, P.; Stark, J.; Swinerd, G. (eds): Spacecraft systems engineering, Fourth Edition (Wiley, Chichester, UK) 2011. ISBN 978-0-470-75012-4

[18] VanOutryve, C. B.: A thermal analysis and design tool for small spacecraft, Master's Thesis, 2008. https://scholarworks.sjsu.edu/etd_theses/3619

HUNGARIAN JOURNAL OF
INDUSTRY AND CHEMISTRY
Vol. 47(1) pp. 33–39 (2019)
hjic.mk.uni-pannon.hu
DOI: 10.33927/hjic-2019-06

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

AUTOMATED LABELING PROCESS FOR UNKNOWN IMAGES IN AN OPEN-WORLD SCENARIO

DÁVID PAPP *[1] AND GÁBOR SZŰCS[1]

[1]Department of Telecommunications and Media Informatics, Budapest University of Technology and Economics, Magyar Tudósok krt. 2., H-1117 Budapest, HUNGARY

Most of the recognition systems presume a controlled, well-defined research setting, where all possible classes that can appear during a test are known a priori. This environment is referred to as the “closed-world” model, while the “open-world” model implies that unknown classes can be incorporated into a recognition algorithm whilst being predicted. Therefore, recognition systems that operate in the real world have to deal with these unknown categories. Our objective was not only to detect data that originate from categories unseen during training, but to identify similarities between pieces of unknown data and then form new classes by automatically labeling them. Our Double Probability Model was extended by an image clustering algorithm, in which Kernel K-means was used. A new procedure, namely the Cluster Classification algorithm for the detection of unknowns and automated labeling, is proposed. These approaches facilitate the transition from open-set recognition to an open-world problem. The Fisher Vector (FV) was used for the mathematical representation of the images and then a Support Vector Machine introduced as a classifier. The measurement of similarity was based on the FV representations. Experiments were conducted on the Caltech101 and Caltech256 datasets of images and the Rand Index was evaluated over the unknown data. The results showed that our proposed Cluster Classification algorithm was able to yield almost the same Rand Index, even though the number of unknown categories increased.

Keywords: open-world problem, cluster classification, image classification, open-set recognition, image clustering

1. INTRODUCTION

In scenarios in the real world, the size of the available dataset continues to increase, therefore, any machine learning algorithm that operates in such an environment has to be capable of preventing growth. This is especially true in the case of image classification, because the growing dataset of tests can pose many difficulties, e.g. it is possible that some of the test images originate from categories that are unseen during training. Recognition systems should detect these unknown images and handle them in an appropriate way. In the rest of the paper the terms "unknown class or category" represent classes or categories that are unseen during training, and “unknown image” denotes images that originate from unknown classes or categories. One way of handling detected unknown images is to measure their similarities and identify new categories. Subsequently, these new categories can be added to the set of known classes. Based on this, three modules are required to solve such problems in the real world, namely a recognition system equipped with an unknown detector, a labeling process and an incremental learning process.

*Correspondence: pappd@tmit.bme.hu

Let us assume that there are K known classes $(C_1, C_2, \dots C_K)$ and U unknown classes in the test set at any given moment, where S_K and S_U denote the sets of known and unknown classes, respectively. Few distinguishable cases depend on the value of U:

1. $U = 0$,
2. $U = 1$,
3. $U > 1$.

Furthermore, a few more cases depend on the amount and type of available information concerning S_U:

(A) Training images,

(B) Set of attributes,

(C) Number of unknown categories (U),

(D) Nothing.

The cases that include 1 or A (e.g. 1A, 2A, 1B, 1C) produce the general multiclass classification because all categories are known a priori and positive-negative samples are available for each category during training. When

$U = 1$, the task is only to identify the unknown images because they originate from the same category, therefore, the similarity measurement is unnecessary. In this paper, the situation when $U > 1$ is considered.

As has been mentioned, 3A represents the traditional multiclass classification. 3B+3C is referred to as transfer learning or zero-shot learning [1], whereas according to the literature the case of 3C+3D is known as open set recognition [2,3] or the open-world problem [4]. The former refers to the detection of images that originate from unknown classes, and the latter includes the detection of unknown images and a labeling process to identify new classes, followed by the incremental learning of these new categories.

Our goal was to tackle the open-world problem as well as develop an algorithm that is able to detect the unknown images and then introduce new classes by automatically labeling the unknown data using unsupervised learning. Previously an algorithm referred to as the Double Probability Model (DPM) [5] was proposed, which is suitable as an unknown detector in an open-set environment.

There are several works that use a variant of Support Vector Machine (SVM) to solve the unknown detection problem, such as the Support Vector Data Description [6], One-class SVM [7, 8], Reject Option SVM (RO-SVM) [9] and the novel Weibull-calibrated SVM (W-SVM) [3]. The latter one was developed to operate under the Compact Abating Probability model, where the probability of class membership decreases (abates) as points move from known data towards unknown space. Scheirer et al. claim that W-SVM outperforms their previous solutions, namely the 1-vs-Set Machine Training algorithm [2] and the Pi-SVM [10]. On the other hand, it was shown that DPM outperforms W-SVM [5], therefore, in this paper the DPM was used for unknown detection. Bendale and Boult defined open world recognition and presented the Nearest Non-Outlier algorithm in [4], which adds object categories incrementally while detecting outliers and managing open space risk. They defined open world recognition in the form of three sequential steps: a multiclass open set recognition function with a novelty detector, a labeling process and an incremental learning algorithm. Although all of these steps should be automated, they presumed labels were obtained by human labeling. The main objective of our work and this paper is to propose an automated labeling process, the so-called Cluster Classification (CC).

In the next section, the DPM and image clustering methods are reviewed, subsequently, a baseline method is suggested for an open-world problem and finally our proposed algorithm, the CC, is presented. The third section contains experimental results and in the last section our conclusion is discussed.

2. Proposed open-world recognition system

2.1 Double Probability Model

The DPM [5] is based on the likelihood of a classifier and can be used with any kind of classifier that provides class membership probabilities for the images. As a result, after training the classifier, it is capable of making predictions with reliability values (scores) for each class, i.e. decision vectors. The range of the scores depends on the type of classifier (sometimes it is from 0 to 1 but it can be over any range. Only one condition is required, namely the larger score for class C_i should represent the higher likelihood of being a member of class C_i. In the training set or a validation set, the instances with corresponding scores are investigated in each class. The ground truth is known for this set, so the positive elements can be selected from each class. In order to calculate the conditional probability that a new instance belongs to class C_i according to its score, the cumulative distribution function (CDF) of positive scores should be determined, therefore, a reverse CDF of negative scores was created:

$$F_{P_i}(x) = p\left(C_i|\text{score} < x\right), \tag{1}$$

$$F_{N_i}(x) = p\left(\neg C_i|\text{score} > x\right), \tag{2}$$

where P_i and N_i denote the positive and negative elements, respectively. Note that the sum of these probabilities is not always equal to 1 (this is not a requirement).

A DPM was constructed based on the CDF and reverse CDF functions. During testing, the focus is on the likelihood of the occurrence of an unknown class compared with any of the known classes. Before the comparison, the probabilities of the known classes should be calculated. Scores (score_i for class C_i) for a new instance are obtained as outputs from the original classifier, and based on them the probability of class C_i occurring can be expressed as described in

$$P_{C_i} = F_{P_i}\left(\text{score}_i\right) \prod_{j=1, j\neq i}^{K} F_{N_j}\left(\text{score}_j\right). \tag{3}$$

An expression for the probability of class C_{K+1} is

$$P_{C_{K+1}} = \prod_{j=1}^{K} F_{N_j}\left(\text{score}_j\right). \tag{4}$$

If the probability of being a member of class C_{K+1} is higher than for any other (known) class, then the new instance will be a member of the unknown class. Otherwise the prediction is based on the original classifier, i.e. the class with the largest score will be selected. The decision with regard to the prediction of test instance j is formalized as

$$d_j = \begin{cases} C_{K+1} & | \quad P_{C_{K+1}} > \max_i \left\{P_{C_i}\right\} \\ \text{argmax}_j \left\{\text{score}_j\right\} & | \quad \text{otherwise} \end{cases} \tag{5}$$

At this point the algorithm is able to make a decision about test data if it originates from an unknown category. Also, should it originate from a known category, then based on the output of the classifier its known category can be determined.

2.2 Unknown image clustering

The image representations were created according to the Bag-of-Words [11, 12] model. Based on their visual content, each image was represented by a single high dimensional vector. In order to create these high-level descriptors, the local attributes of the images were investigated by calculating the low-level Scale Invariant Feature Transform (SIFT) [13] descriptor. Next, the Gaussian Mixture Model (GMM) [14–16] was used to define the visual code words and the Fisher Vectors [17, 18] to encode the low-level descriptors into high-level descriptors based on the visual code words. The Fisher Vectors were the final representations (image descriptors) of the images and were used as the input data for the clustering algorithm. After the final clusters of Fisher Vectors were formed, the image clusters could be produced by substituting the Fisher Vectors for the corresponding images.

The basis of our clustering approach is the well-known K-means clustering algorithm [19] which consists of two important inputs, namely the initial cluster centers and the number of clusters. The K-means clustering algorithm aims to minimize the sum of squared distances from all points to their cluster centers:

$$E = \min \left(\sum_{l=1}^{k} \sum_{x_i \in C_l} \|x_i - z_l\|^2 \right), \tag{6}$$

where k denotes the number of clusters, x_i represents a member of cluster C_l and z_l stands for the center of it.

However, the Fisher Vector consists of 65,791 dimensions, and the basic K-means clustering algorithm performs less efficiently when the clusters are non-linearly separable or the data contains arbitrarily shaped clusters of different densities. Therefore, an upgraded version of the K-means clustering algorithm was applied in the recognition system referred to as Kernel K-means [20–22]. The objective function of Kernel K-means is still to minimize the sum of squared distances, but it uses the kernel trick to transform the data points into infinite feature space $x_i \to \vartheta(x_i)$, as can be seen in

$$E = \min \left(\sum_{l=1}^{k} \sum_{x_i \in C_l} \left\| \vartheta(x_i) - \frac{\sum_{x_j \in C_l} \vartheta(x_j)}{N_l} \right\|^2 \right), \tag{7}$$

where N_l denotes the number of images in cluster C_l. The trick here is that explicit calculations in the feature space are never required, since transformed data points are only present as part of an inner product. Therefore, they can be substituted for their kernel representatives (the Gaussian kernel was implemented here).

In order to reduce the randomness of final clusters, the PlusPlus cluster center initialization algorithm was used before the iterative steps, which was proposed by D. Arthur and S. Vassilvitskii [23]. This approach aims to spread out the initial cluster centers and accelerate their convergence. The first cluster center is randomly selected from the data points, after that each subsequent cluster center is chosen from the data points with a probability proportional to its squared distance from the closest existing cluster center.

In the following sub-sections, the usage of the presented methods is discussed.

2.3 Baseline method

In this section, a baseline method of open world recognition is presented. First, at training time the classifier of the training data is trained with K known classes, then, at testing time classification of the test data ($K+U$ classes) is performed. The DPM is applied to the output of the classifier to detect unknown images U^{DPM}:

$$U^{\mathrm{DPM}} = \bigcup_{j=1}^{N_U} \{I_j | d_j = C_{K+1}\} \tag{8}$$

where I_j represents test instance j, N_U denotes the number of test instances in the test data, d_j stands for the decision of the DPM, and $\bigcup \{\dots\}$ is the operation of union.

Now, let us assume that information concerning U was provided (as in the case 3C), and U was used as the number of clusters. The Kernel K-means PlusPlus cluster center initialization algorithm (KK^{++}) was performed on U^{DPM} with $k = U$ clusters (which is the input parameter for the KK^{++}), and then the appropriate labels were assigned to the unknown images:

$$\begin{gathered} L_j = C_{K+i} | i = \mathrm{out}\left(\mathrm{KK}^{++}\right) \\ j = 1 \dots M, \quad i = 1 \dots U \end{gathered} \tag{9}$$

where M represents the number of unknown images; L_j and C_i denote the label of unknown image U_j^{DPM} and cluster identity, respectively. This concludes the baseline method for automated labeling. At this point the classifier can be retrained based on the previously known and new labels, and then the new test data classified.

2.4 Cluster Classification

In this section our proposed CC approach is presented, which is suitable for unknown detection and automated labeling. This algorithm contains extended training and testing phases. In training time, a classifier of the training data is trained with K known classes, then a pseudo-cluster is also created based on the K known categories. This means that the ground truth class labels are implemented rather than a clustering algorithm (to determine the final clusters), i.e. each category is a cluster. Subsequently, the images are substituted for their Fisher Vector

representations and the cluster centers calculated which will be used in the testing phase.

Let us assume T categories are found in the testing phase, and that $T > K$. The test data is classified into the K known categories and a DPM applied based on the decision vectors to detect the unknown images U^{DPM}. The next step is to form clusters using the Kernel K-means clustering algorithm starting from the K cluster centers that were calculated at training time from the pseudo-cluster. Afterwards, the remaining T-K cluster centers are determined following the PlusPlus initiation protocol. Furthermore, the training and test datasets were used together as the input data. Basically, with these modifications it was possible to guide the clustering algorithm, therefore, create more accurate clusters.

The following step of the testing phase is to classify the clusters $\{C_i\}$ by weighted majority voting of the members of the cluster. The vote is based on the class membership probabilities $(P_{C_i}; i = 1 \dots K+1)$ calculated in Eqs. 3–4. As was seen in Section 1, the definition of problem 3C assumes that the number of unknown categories exceeds 1. Nonetheless, the output of the DPM only yields $K+1$ alternatives instead of T. In spite of this, the classification of clusters that depend on $\{P_{C_i}\}$ can increase the number of alternatives to T as will be seen later. In Section 1, a differentiation was made between known and unknown images, and now this differentiation is broken down even more. The training data contains only known images, because each of them belongs to one of the set of known categories (S_K). From now on, the union of known images of the training data will be denoted by K^{GT} as can be seen in:

$$K^{\mathrm{GT}} = \bigcup_{j=1}^{N_K} \{I_j\} \tag{10}$$

where N_K stands for the number of images in the training data. On the other hand, the test data contains both known and unknown images. Furthermore, based on the output of DPM, the test data can be divided into two different subsets, namely predicted known images (K^{DPM}) and predicted unknown images (U^{DPM}), as can be seen in Eqs. 11 and 8, respectively.

$$K^{\mathrm{DPM}} = \bigcup_{j=1}^{N_U} \{I_j | d_j \neq C_{K+1}\} \tag{11}$$

The weight of the images can be calculated based on the cluster coherence. The coherence of a cluster can be determined by comparing the number of known images to the number of predicted unknown images inside that given cluster. It should be noted that known images inside the clusters either originate from K^{GT} or K^{DPM}, while the predicted unknown images are all part of U^{DPM}. If the number of known images exceeds the number of unknown images it is implied that a cluster exhibits "known coherence" (KC), and "unknown coherence" (UC) vice versa, as described in:

$$C^{\mathrm{coh}} = \begin{cases} \mathrm{KC} \mid \left\|\{K^{\mathrm{GT}} \cup K^{\mathrm{DPM}}\}\right\| \geq \left\|U^{\mathrm{DPM}}\right\| \\ \mathrm{UC} \mid \left\|\{K^{\mathrm{GT}} \cup K^{\mathrm{DPM}}\}\right\| < \left\|U^{\mathrm{DPM}}\right\| \end{cases} \tag{12}$$

where $\|X\|$ represents the number of elements in X, and the superscript coh indicates the coherence of cluster C.

The weights can be calculated as described in Eqs. 13 and 14. Intuitively, if an image is known and located inside cluster UC, then it is "punished" by assigning a lower weight to it; and vice versa, an unknown image is given a lower weight inside cluster KC. Moreover, the larger the difference between the numbers of known and unknown images implies a more severe punishment with regard to the value of weights.

$$w_j^{\mathrm{KC}} = \begin{cases} 1 + \frac{(\#\mathrm{known} - \#\mathrm{unknown})}{(\#\mathrm{known} + \#\mathrm{unknown})} \mid I_j \notin U^{\mathrm{DPM}} \\ 1 - \frac{(\#\mathrm{known} - \#\mathrm{unknown})}{(\#\mathrm{known} + \#\mathrm{unknown})} \mid I_j \in U^{\mathrm{DPM}} \end{cases} \tag{13}$$

$$w_j^{\mathrm{UC}} = \begin{cases} 1 + \frac{(\#\mathrm{known} - \#\mathrm{unknown})}{(\#\mathrm{known} + \#\mathrm{unknown})} \mid I_j \in U^{\mathrm{DPM}} \\ 1 - \frac{(\#\mathrm{known} - \#\mathrm{unknown})}{(\#\mathrm{known} + \#\mathrm{unknown})} \mid I_j \notin U^{\mathrm{DPM}} \end{cases} \tag{14}$$

Thereafter the final decision vector of cluster C_i can be calculated as:

$$V_i = \frac{1}{N_i} \sum_{j=1}^{N_i} w_j \times d_j \tag{15}$$

where N_i denotes the number of images in cluster C_i, w_j represents the weight and d_j stands for the decision vector ($\{P_{C_i}\}$) of image j. Note that d_j possesses $K+1$ elements ($+1$ from DPM), therefore, vector V_i also possesses $K+1$ elements. Consequently, the element with the maximum value of V_i determines the category of cluster C_i. The classification of cluster C_i is formalized in:

$$D_i = \begin{cases} \text{new class} & \mid \quad V_{K+1} = \max_j\{V_j\} \\ \mathrm{argmax}_i\{V_i\} & \mid \quad otherwise \end{cases} \tag{16}$$

The results of the classification of the clusters can be considered as a labeling proposal, i.e. label each image inside cluster C_i according to D_i. When decision D_i for cluster C_i is that it is part of a known category, then each image inside C_i gets labeled with the same category. On the other hand, when $D_i =$ a new class, a new category is created and each image in C_i gets labeled with the new category. Basically, the CC algorithm follows this labeling proposal.

3. Experimental Results

In order to measure the efficiency of the labeling process, experiments were conducted on the Caltech101 [24] and

Figure 1: Example images from the Caltech101 and Caltech256 datasets. The airplane, butterfly and windmill categories are represented by the left, middle and right columns, respectively.

Caltech256 [25] datasets. Example images from these datasets are shown in Fig. 1.

The former consists of 101 categories and 8,677 images, while the latter is composed of 30,607 images from 256 different classes. To create an open-world environment, 50 known and 50 unknown categories were randomly selected from the Caltech101 dataset, and 100 of both categories from the Caltech256 dataset. These random selections were repeated 5 times in order to calculate the average of the results of each experiment to obtain a more comprehensive overview of the efficiency of the CC algorithm with regard to these datasets. All of the known categories were available from the beginning of the tests, but the unknown categories were added incrementally over 10 steps, and in each step the Rand Index (RI),

$$\mathrm{RI} = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}}, \tag{17}$$

was evaluated over the unknown images, where TP, TN, FP, and FN denote the number of true positive, true negative, false positive and false negative decisions, respectively. The RI measures the similarity between the ground truth and predicted labels of the unknown images, in other words, the percentage of correct decisions.

Two methods were assessed and compared, namely the baseline method (DPM+KK) and the CC, which were discussed in Section 2.3 and 2.4, respectively. Both procedures used Fisher Vectors to mathematically represent the images encoded from 128 dimensional SIFT descriptors using a GMM consisting of 256 code words; a SVM equipped with a radial basis function (RBF) kernel was applied as a classifier. The results can be seen in Fig. 2 and Table 1.

The first diagram shows the results obtained from the Caltech101 dataset and the second from the Caltech256 dataset. The DPM with Kernel K-means and the CC are represented by dashed and solid lines, respectively. In both experiments, the CC algorithm yielded a higher RI, although during the first step the difference between the two methods was minimal. It can be seen that the RI of DPM+KK starts to decrease as the number of unknown categories increases, while the CC remains by and large unchanged.

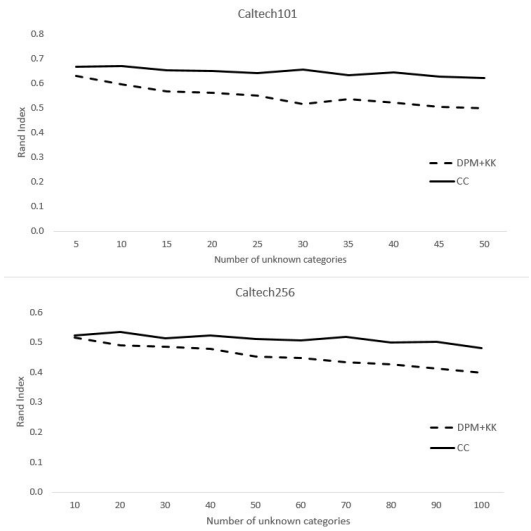


Figure 2: Averaged results of the 5-5 different test datasets that were randomly selected from the Caltech101 and Caltech256 datasets. The RI is plotted against the number of unknown categories. The diagrams compare the labeling performance of the DPM with Kernel K-means (DPM+KK) against the CC.

Table 1: Summary of the results obtained from the test data with the baseline (DPM+KK) and CC methods using the Caltech101 and Caltech256 datasets. The baseline column contains the RI values evaluated which depend on the number of unknown categories (un. cat.), and the CC column presents the improvements that result from CC as a percentage.

Caltech101			Caltech256		
un. cat.	base-line	CC (%)	un. cat.	base-line	CC (%)
5	0.629	6	10	0.514	1
10	0.594	13	20	0.489	9
15	0.567	15	30	0.484	6
20	0.561	16	40	0.478	9
25	0.550	17	50	0.452	13
30	0.514	28	60	0.448	13
35	0.536	18	70	0.433	19
40	0.522	23	80	0.426	17
45	0.505	24	90	0.412	22
50	0.497	25	100	0.397	21

4. Conclusion

In this paper, the problem of open world recognition was reviewed and the possible cases were differentiated based on our prior knowledge and actual information about the test data and, thus, the unknown space. The DPM and Kernel K-means algorithm were also reviewed in brief, followed by the presentation of two approaches, which perform multi-class classification, automatically detect unknown images and propose a labeling for them. The first method is a baseline technique where DPM was sequentially applied followed by Kernel K-means with a PlusPlus cluster center initialization algorithm. However, our proposed CC is a complex method of combining the unknown detector and clustering algorithm that seeks to determine the identity of formed clusters, while refining the decisions made by the classifier and unknown detector. The CC algorithm constructs a specific weight system to reward or punish images which were placed into a category that is presumably unsuitable for their estimated identity. Multiple experiments were conducted on two large datasets (Caltech101 and Caltech256), and the RI evaluated with regard to the unknown images. The results showed that the CC outperformed the baseline method, and was able to maintain almost the same RI, while the number of unknown categories increased.

Acknowledgement

The research was supported by the ÚNKP-18-3 New National Excellence Program of the Ministry of Human Capacities.

REFERENCES

[1] Lampert, C. H.; Nickisch, H.; Harmeling, S.: Learning to detect unseen object classes by between-class attribute transfer, *2009 IEEE Conference on Computer Vision and Pattern Recognition,* 2009, pp. 951–958 ISBN: 978-1-4244-3992-8 DOI: 10.1109/CVPR.2009.5206594

[2] Scheirer, W. J.; de Rezende Rocha, A.; Sapkota, A.; Boult, T. E.: Toward open set recognition, *IEEE T. Pattern Anal.,* 2013 **35**(7), 1757–1772 DOI: 10.1109/TPAMI.2012.256

[3] Scheirer, W. J.; Jain, L. P.; Boult, T. E.: Probability models for open set recognition, *IEEE T. Pattern Anal.,* 2014 **36**(11), 2317–2324 DOI: 10.1109/TPAMI.2014.2321392

[4] Bendale, A.; Boult, T.: Towards open world recognition, *2015 IEEE Conference on Computer Vision and Pattern Recognition,* 2015, pp. 1893–1902 ISBN: 978-1-4673-6964-0 DOI: 10.1109/CVPR.2015.7298799

[5] Papp, D.; Szűcs, G.: Double probability model for open set problem at image classification, *INFORMATICA,* 2018 **29**(2), 353–369 DOI: 10.15388/Informatica.2018.171

[6] Tax, D. M.; Duin, R. P.: Support vector data description, *Machine Learning,* 2004 **54**(1), 45–66 DOI: 10.1023/B:MACH.0000008084.60811.49

[7] Cevikalp, H.; Triggs, B.: Efficient object detection using cascades of nearest convex model classifiers, *2012 IEEE Conference on Computer Vision and Pattern Recognition,* 2012, pp. 3138–3145 ISBN: 978-1-4673-1226-4 DOI: 10.1109/CVPR.2012.6248047

[8] Schölkopf, B.; Platt, J. C.; Shawe-Taylor, J.; Smola, A. J.; Williamson, R. C.: Estimating the support of a high-dimensional distribution, *Neural Comput.*, 2001 **13**(7), 1443–1471 DOI: 10.1162/089976601750264965

[9] Zhang, R.; Metaxas, D. N.: RO-SVM: Support vector machine with reject option for image categorization, In: Chantler, M.; Fisher, B.; Trucco, M.; (Eds.): Proceedings of the British Machine Conference (BMVA Press, UK) 2006, pp. 123.1-123.10. ISBN: 1-901725-32-4 DOI: 10.5244/C.20.123

[10] Jain, L. P.; Scheirer, W. J.; Boult, T. E.: Multi-class open set recognition using probability of inclusion, In: Fleet, D.; Pajdla, T.; Schiele, B.; Tuytelaars, T.; (Eds.): Computer Vision – ECCV 2014., ECCV 2014. Lecture Notes in Computer Science, **8691** (Springer, Cham, Switzerland) 2014, pp. 393–409. ISBN: 978-3-319-10577-2 DOI: 10.1007/978-3-319-10578-9_26

[11] Fei-Fei, L.; Fergus, R.; Torralba, A.: Recognizing and learning object categories, *2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Short course, 2007 http://people.csail.mit.edu/torralba/shortCourseRLOC/

[12] Lazebnik, S.; Schmid, C.; Ponce, J.: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories, *2006 IEEE Conference on Computer Vision and Pattern Recognition,* 2006 **2**, pp. 2169–2178 ISBN: 0-7695-2597-0 DOI: 10.1109/CVPR.2006.68

[13] Lowe, D. G.: Distinctive image features from scale-invariant keypoints, *Int. J. Comput. Vision,* 2004 **60**(2), 91–110 DOI: 10.1023/B:VISI.0000029664.99615.94

[14] Reynolds, D. A.: Gaussian mixture models, In: Li, S. Z.; (Ed.): Encyclopedia of Biometric Recognition, 1st ed., (Springer, Boston, USA) 2009, pp. 659–663 ISBN: 978-0-387-73003-5 DOI: 10.1007/978-1-4899-7488-4_196

[15] Tomasi, C.: Estimating Gaussian mixture densities with EM: A tutorial (Tech. rep., Duke University) 2004 https://www2.cs.duke.edu/courses/spring04/cps196.1/handouts/EM/tomasiEM.pdf

[16] Browne, R. P.; McNicholas, P. D.; Sparling, M. D.: Model-based learning using a mixture of mixtures of Gaussian and uniform distributions, *IEEE T. Pattern Anal.,* 2012 **34**(4), 814–817 DOI: 10.1109/TPAMI.2011.199

[17] Perronnin, F.; Dance, C.: Fisher kernel on visual vocabularies for image categorization, *2007 IEEE Conference on Computer Vision and Pattern*

Recognition, 2007, pp. 1–8 ISBN: 1-4244-1179-3 DOI: 10.1109/CVPR.2007.383266

[18] Perronnin, F.; Sánchez, J.; Mensink, T.: Improving the Fisher kernel for large-scale image classification, In: Daniilidis, K; Maragos, P.; Paragios, N.; (Eds.): Computer Vision – ECCV 2010., ECCV 2010. Lecture Notes in Computer Science, **6314** (Springer, Berlin, Germany) 2010, pp. 143–156 ISBN: 978-3-642-15560-4 DOI: 10.1007/978-3-642-15561-1_11

[19] MacQueen, J.: Some methods for classification and analysis of multivariate observations, In: Le Cam, L. M.; Neyman, J.; (Eds.): Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, **1** (University of California Press, Berkeley, USA) 1967 pp. 281–297

[20] Chitta, R.; Jin, R.; Havens, T. C.; Jain, A. K.: Approximate kernel k-means: Solution to large scale kernel clustering, In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (ACM, New York, USA) 2011, pp. 895–903 ISBN: 978-1-4503-0813-7 DOI: 10.1145/2020408.2020558

[21] Dhillon, I. S.; Guan, Y.; Kulis, B.; Kernel k-means: spectral clustering, and normalized cuts, In: Proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (ACM, New York, USA) 2004, pp. 551–556 ISBN: 1-58113-888-1 DOI: 10.1145/1014052.1014118

[22] Papp, D.; Szűcs, G.: MMKK++ algorithm for clustering heterogeneous images into an unknown number of clusters, *ELCVIA: Electronic Letters on Computer Vision and Image Analysis,* 2017 **16**(3), 30–45 DOI: 10.5565/rev/elcvia.1054

[23] Arthur, D.; Vassilvitskii, S.: k-means++: The advantages of careful seeding, In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, (Society for Industrial and Applied Mathematics, Philadelphia, USA) 2007, pp. 1027–1035 ISBN: 978-0-898716-24-5

[24] Fei-Fei, L.; Fergus, R.; Perona, P.: Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories, *Comput. Vis. Image Und.,* 2007 **106**(1), 59–70. DOI: 10.1016/j.cviu.2005.09.012

[25] Griffin, G.; Holub, A.; Perona, P.: The Caltech 256, *Caltech, Tech. Rep.*, 2012

HUNGARIAN JOURNAL OF
INDUSTRY AND CHEMISTRY
Vol. 47(1) pp. 41–48 (2019)
hjic.mk.uni-pannon.hu
DOI: 10.33927/hjic-2019-07

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

MODEL REFERENCE ADAPTIVE CONTROL FOR TELEMANIPULATION

NÁNDOR FINK *[1]

[1]Department of Mechatronics, Optics and Engineering Informatics, Budapest University of Technology and Economics, Bertalan Lajos utca 4-6, Budapest, 1111, HUNGARY

A 1-DOF (degree-of-freedom) telemanipulation system is presented in this paper. The paper focuses on disturbance compensation of the haptic force feedback. The master and slave devices are connected via serial ports. The mechanism, which is applied as a human interface device, is subject to perceptible internal friction that must be eliminated. As a result, the operator would only feel the force feedback from the manipulated environment. The main contribution of this paper is the presentation of the telemanipulation device with a model reference adaptive control that compensates for the friction force using a direct model-based sliding mode algorithm.

Keywords: telemanipulation, friction compensation, sliding mode control

1. INTRODUCTION

In connection with the rapid spread of the Internet over the past few years, research about CSCW (Computer-Supported Cooperative Work) technology is being vehemently conducted. It is expected that this network-based technology will make collaboration in connection with human intellectual activities between distantly connected people easier. In the field of CSCW technology, the data to be processed are only images, sounds and other data for computers, little attention has been paid to multi-modal collaboration including physical contact. Our focus concerns networked multi-modal collaboration especially between those including haptics.

A collaboration tool that facilitates a better connection between laboratories, offices and factories in the manufacturing industry has become necessary [1–3].

Research into the bilateral control of master-slave manipulators was conducted in 1992. In [4] a 1-DOF telemanipulation system was tested. The research covered the dynamics of the human operator as well as the device. Three levels of ideal responses were determined, in all cases a force signal to the master arm was the input. The position of the two arms was identical in the first level, whereas the force response of the two arms was identical in the second, and both responses were identical in the third.

In [5] a four-channel control architecture was examined. Four models were tested on a teleoperation system, namely on an admittance-admittance, impedance-admittance, admittance-impedance and impedance-impedance pair. Based on the types of model, a one-one suggestion was made for the control architecture in each case, namely Position-Position, Position-Force, Force-Position and Force-Force.

In addition to the basic criteria of stability, the ease of usability becomes increasingly important. To ensure comfortable and ergonomic telemanipulation, the operator should control the slave device more smoothly and precisely. Furthermore, the elimination of perturbations in all circumstances is necessary. For this purpose, the design of a model-reference adaptive control with a sliding mode friction compensator is promising.

2. Theoretical background of the experiment

2.1 *Sliding mode-based disturbance elimination*

The main concept is the design of a reduced-order state observer for a partially perturbed linear system with infinite gain. The traditional roles of the system and observer are exchanged. The system is forced to follow the states of the unperturbed ideal model. Infinite gain is ensured by a sliding mode.

Consider the following partially perturbed linear system that consists of external disturbances and uncertain parameters which satisfy the so-called Drazenovic condition, written in the regular form of a state equation,

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix} = \begin{bmatrix}\overline{\mathbf{A}}_{11} & \overline{\mathbf{A}}_{12}\\ \overline{\mathbf{A}}_{21}+\Delta\mathbf{A}_{21} & \overline{\mathbf{A}}_{22}+\Delta\mathbf{A}_{22}\end{bmatrix}\begin{bmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{bmatrix} + \begin{bmatrix}0\\ \overline{\mathbf{B}}_2+\Delta\mathbf{B}_2\end{bmatrix}\mathbf{u}^0 + \begin{bmatrix}0\\ \mathbf{E}_2\end{bmatrix} f(t) \quad (1)$$

*Correspondence: finknandor@mogi.bme.hu

where $\mathbf{x}_1 \in \mathbb{R}^{n-m}$ denotes the vector of non-perturbed state variables, $\mathbf{x}_2 \in \mathbb{R}^m$ stands for the vector of perturbed state variables, $\mathbf{u}^0 \in \mathbb{R}^m$ represents the input of the real system, $\overline{\mathbf{A}}_{ij}(i,j=1,2)$ and $\overline{\mathbf{B}}_2$ are the nominal or desired (ideal) matrices of the system, respectively, $\Delta\mathbf{A}_{2j}(j=1,2)$ and $\Delta\mathbf{B}_2$ denote the bounded parameter perturbations, and $f(t)$ stands for the bounded external disturbance. The perturbed state variables are estimated by a discontinuous observer [6]

$$\frac{\mathrm{d}}{\mathrm{d}t}\widehat{\mathbf{x}}_2 = \overline{\mathbf{A}}_{21}\mathbf{x}_1 + \overline{\mathbf{A}}_{22}\widehat{\mathbf{x}}_2 + \overline{\mathbf{B}}_2\left(\mathbf{u}^0 + \boldsymbol{\nu}\right) \tag{2}$$

where $\boldsymbol{\nu}$ denotes the discontinuous term. Let us design the following sliding surface:

$$\boldsymbol{\sigma} = \begin{bmatrix}\mathbf{I} & -\mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{x}_2 \\ \widehat{\mathbf{x}}_2\end{bmatrix} = 0 \tag{3}$$

where $\mathbf{I} \in \mathbb{R}^{m\times m}$ represents the identity matrix and $\boldsymbol{\sigma} \in \mathbb{R}^m$ stands for the distance from the surface. $\boldsymbol{\sigma}$ must tend to zero. Let us calculate the elements of the discontinuous term $\boldsymbol{\nu}$ in following way:

$$\nu_i = G_i\,\mathrm{sign}(\sigma_i), \tag{4}$$

where G_i is the gain of the sliding mode controller. The implementation of a sliding mode means that the signs of $\sigma_i = 0$ and ν_i change at an infinitely high frequency [7]. $\boldsymbol{\nu}$ can be substituted by its mean value denoted by $\boldsymbol{\nu}_{\mathrm{eq}}$. By comparing the second line of Eq. 1 and Eq. 2:

$$\left(\overline{\mathbf{B}}_2 + \Delta\mathbf{B}_2\right)\boldsymbol{\nu} = \Delta\mathbf{A}_{21}\mathbf{x}_1 + \left(\overline{\mathbf{A}}_{22} + \Delta\mathbf{A}_{22}\right)\mathbf{x}_2 - \overline{\mathbf{A}}_{22}\widehat{\mathbf{x}}_2 + \Delta\mathbf{B}_2\mathbf{u}^0 + \mathbf{E}_2 f(t) \tag{5}$$

According to Eq. 5, $\boldsymbol{\nu}$ can be used to estimate the perturbation. As a result, the response of the perturbed system in terms of $\mathbf{u}-\boldsymbol{\nu}$ will be identical to that of the unperturbed ideal system, $\mathbf{u}$.

The main problem concerning the sliding mode is the chattering caused by the infinitely alternating frequency of $\boldsymbol{\nu}$. To avoid uncontrolled resonances of the unmodeled dynamics of the real system, $\boldsymbol{\nu}$ is substituted by $\boldsymbol{\nu}_{\mathrm{eq}}$ (the continuous equivalent of $\boldsymbol{\nu}$. In practice, it is impossible to calculate the equivalent control $\boldsymbol{\nu}_{\mathrm{eq}}$ precisely, but it can be estimated by a low-pass filter for $\boldsymbol{\nu}$ as shown in Fig. 1, where two loops can be seen.

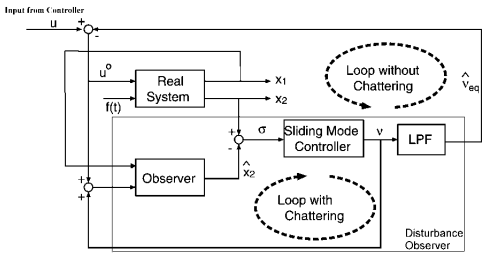


Figure 1: Sliding mode-based disturbance compensation.

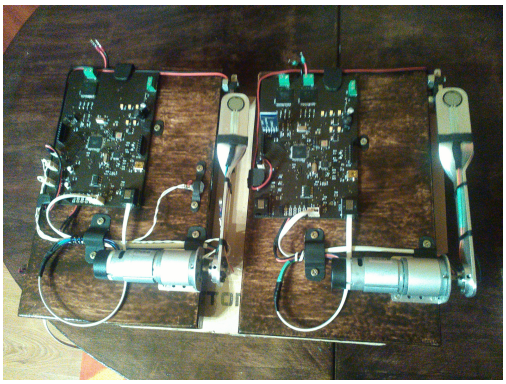


Figure 2: The haptic device.

The observer – sliding-mode control loop is calculated by the computer and should be as fast as possible to achieve an ideal sliding mode [8, 9]. Since a reduced order observer is used, $\mathbf{x}_2$ of the real system is measured in terms of the disturbance compensation. Of course, all state variables might be measured in the outer control loop. $\widehat{\boldsymbol{\nu}}_{\mathrm{eq}}$, the estimation of $\boldsymbol{\nu}_{\mathrm{eq}}$, is added to the control signal of the outer control loop.

3. Application

3.1 Tuning the disturbance compensation

The master device consists of a DC motor and an arm (Fig. 2). The slave device is identical to the master device. Their roles are interchangeable. The arm is not rigid since the force is measured by strain gauges. In the case of the ideal telemanipulation model, the master device is not subject to friction nor mass (inertia). Of course it is impossible to construct an ideal master device. The goal of the sliding mode-based disturbance compensation is to force the master device to follow a model subject to significantly reduced degrees of friction and inertia. The model of the master and slave devices is shown in Fig. 3. The position of the motor is controlled by a simple Proportional Derivative (PD) controller. The results of three simulations were compared:

- PD controller with the parameters of the real motor
- PD controller with the parameters of the desired motor
- PD controller with the parameters of the real motor which were modified by the addition of sliding-mode disturbance compensation.

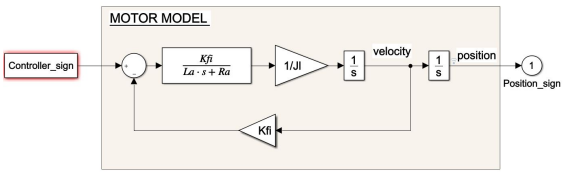


Figure 3: Model of the master and slave devices.

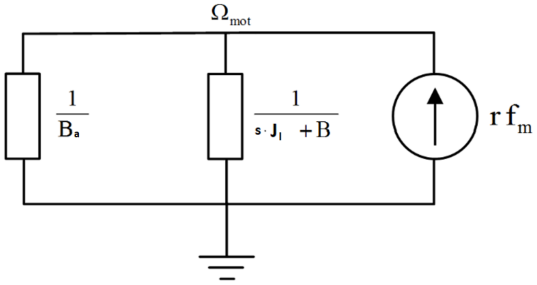


Figure 4: Model of the feedback force.

The main goal of this chapter is to tune the disturbance compensation component. If the results of the last two simulations are similar, sliding-mode disturbance compensation can be used in a bilateral telemanipulation system.

The parameters used in the model are calculable. A 1-DOF telemanipulation system was studied where the effect of the feedback force of the master arm was modelled [10] as follows (Fig. 4):

$$\frac{\mathbf{\Omega}_{\mathrm{mot}}}{\mathbf{r}\cdot\mathbf{f}_m} = \frac{1}{s\cdot J_l + B + B_a} \tag{6}$$

$$\frac{\Omega_{\mathrm{mot}}}{f_m} = \frac{\frac{r}{B+B_a}}{s\cdot\frac{J_l}{B+B_a}+1} = \frac{\frac{r}{B_m}}{s\cdot T+1}, \tag{7}$$

where s is a complex variable and T denotes the time constant of the manipulator:

$$T = \frac{J_{\mathrm{l}}}{B_m} \tag{8}$$

J_{l} is calculated according to:

$$J_{\mathrm{l}} = J_{\mathrm{r}} + J_{\mathrm{tr2}} + J_{\mathrm{a}}, \tag{9}$$

and represents the resultant moment of inertia, J_{r} stands for the inertia of the rotor, J_{tr2} is the transmission inertia and J_{a} denotes the inertia of the arm.

$$\mathbf{B}_{\mathrm{m}} = B + B_{\mathrm{a}} \tag{10}$$

$\mathbf{B}_{\mathrm{m}}$ represents the resultant damping, B the mechanical damping and B_{a} the damping caused by the armature resistance.

$$B_{\mathrm{a}} = \frac{K_{\mathrm{fi}}^2}{R_{\mathrm{a}}} \tag{11}$$

K_{fi} denotes the torque constant of the motor and R_{a} its resistance. Ω_{mot} stands for the angular speed of the motor, f_{m} the force- and $\mathbf{r}$ the vector of the lever arm with regard to the torque of the motor.

In this experiment, the following constants were used for the aforementioned parameters: $K_{\mathrm{fi}} = 0.1222$ [Nm/A], $R_{\mathrm{a}} = 80$ [Ω], $L_{\mathrm{a}} = 0.0011$ [H] and $J_{\mathrm{l}} = 0.12$ [kg m^2].

The model of the ideal motor took friction into account (Fig. 5), however, the other models accounted for this separately.

3.2 Design of the disturbance compensation component

In terms of human sensation, the electrical time constant of the system (the inductance of the DC motor) is negligible. Two state variables are used: the position of the arm φ and the angular speed ω. The system equation must be written in the same form as (Eq. 1) where:

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\varphi\\ \omega\end{bmatrix} = \begin{bmatrix}0 & 1\\ 0 & a_{\mathrm{motor}}\end{bmatrix}\begin{bmatrix}\varphi\\ \omega\end{bmatrix} + \begin{bmatrix}0\\ b_{\mathrm{motor}}\end{bmatrix}\mathbf{u}^0, \tag{12}$$

a_{motor} and b_{motor} are the two perturbed parameters. No external disturbance is applied. The observer is designed for the state variable ω.

It should be noted that in the case of the ideal system, friction is accounted for in the model. The ideal parameters were selected in terms of friction and inertia. It is important to choose the ideal parameters wisely because these two parameters will determine the trajectory followed by the position signal of the compensated motor. If the values of the ideal parameters are too unrealistic, the compensated signal would not be able to follow the trajectory! In this simulation, the following parameters were used:

$$J_{\mathrm{l\_i}} = \frac{J_{\mathrm{l}}}{12} \quad B_{\mathrm{m\_i}} = \frac{B_{\mathrm{m}}}{5} \tag{13}$$

where $J_{\mathrm{l\_i}}$ denotes the inertia with regard to the model of the ideal motor and $B_{\mathrm{m\_i}}$ stands for the coefficient of friction concerning the model of the ideal motor.

The compensation component consists of the observer with the following transfer function:

$$\frac{b_{\mathrm{motor}}}{s + a_{\mathrm{motor}}} \tag{14}$$

where a_{motor} is calculated by:

$$a_{\mathrm{motor}} = \frac{B_{\mathrm{m\_i}}}{J_{\mathrm{l\_i}}} - \frac{K_{\mathrm{fi}}^2}{R_{\mathrm{a}}\cdot J_{\mathrm{l\_i}}} \tag{15}$$

and b_{motor} by:

$$b_{\mathrm{motor}} = \frac{K_{\mathrm{fi}}}{R_{\mathrm{a}}\cdot J_{\mathrm{l\_i}}} \tag{16}$$

Then a PD position controller is added to each model of a motor (Fig. 5) and the component concerning the sliding-mode compensator connected to the first real model of a motor (Fig. 6).

An estimated value of the velocity is specified in this section which will be compared with that of the actual velocity. This will be the input for the component of the sliding-mode controller that consists of a signum function and a gain block. From here a positive feedback was applied from the output of the sliding-mode controller to the input of the observer, and the output of the sliding-mode controller filtered by a 3$^{\mathrm{rd}}$ order low-pass filter as the output of the whole compensation component (Fig. 6).

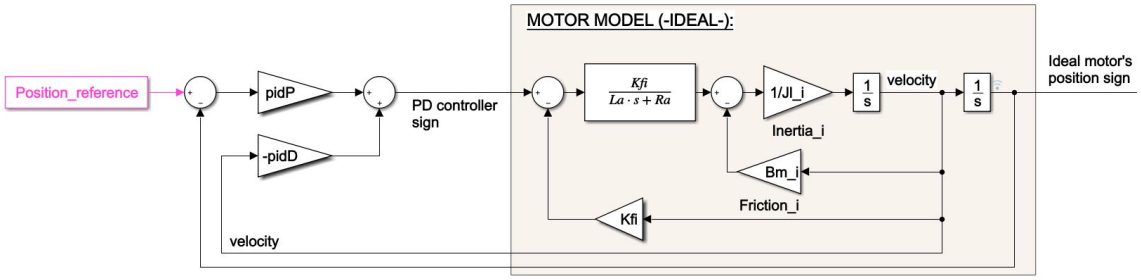


Figure 5: Model of the ideal motor with the application of a PD controller.

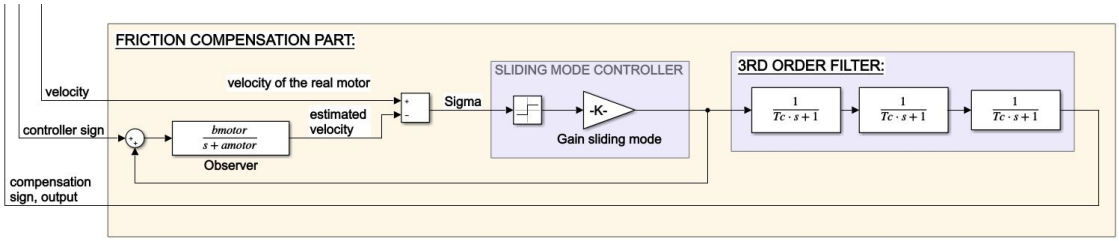


Figure 6: The component with regard to the compensation of friction.

3.3 Calibration of the model

To specify the parameters of the system, first the aforementioned model must be implemented. In the first case, a position controller was used for tuning (Fig. 7). It is worth noting that the position signal of the real motor reaches its final value by oscillation and overshooting, while the graphs of the ideal motor and the compensated motor are smoother. It has to be mentioned that the PD controller is not optimised for a stand-alone task, rather for the compensation. In this experiment the following constants were used for the aforementioned parameters: $\mathrm{pidP} = 190,\ \mathrm{pidP} = 60$ and $\mathrm{Gain}_{\mathrm{sliding}} = 30,000$.

In the case of a 1-second-long simulation, the difference between the position trajectory of the ideal and compensated motors is visible (Fig. 8).

In Fig. 9 it is visible that the sliding-mode controller is inactive for about 0.2 seconds which is the time necessary for the controller to reach the sliding surface.

Between -0.04 and $+0.04$ the active sliding-mode becomes visible (Fig. 10). There is a noticeable level of chattering originating from the output signal of the sliding-mode controller (Fig. 11). This was eliminated by the 3$^{\mathrm{rd}}$ order low-pass filter (Fig. 12).

3.4 Verification of the model

Subsequently, the calibration component follows on from verification of the model. The former simulations were run under nearly ideal conditions, but now a relevant Coulomb friction will be added to the system as well as an additional step function load. In Fig. 13 the supplemented system is shown only for the model of the motor that compensated for friction using real parameters (the models of the ideal and simple real motors were excluded to save space).

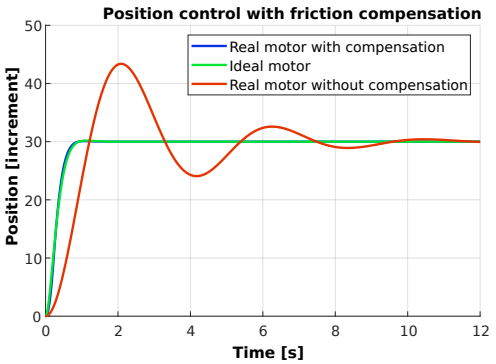


Figure 7: The compensated signal follows the stated trajectory.

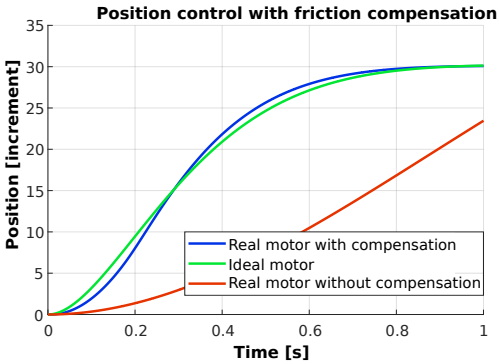


Figure 8: The difference between the position signals of the ideal and compensated motors.

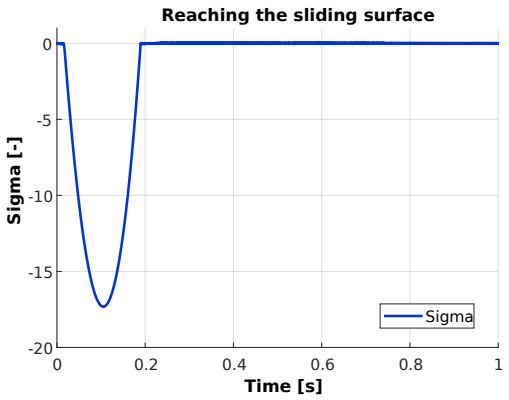


Figure 9: Reaching the sliding surface.

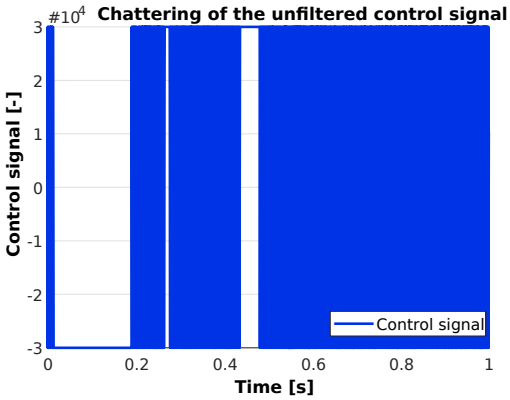


Figure 11: Chattering originating from the output of the sliding-mode controller.

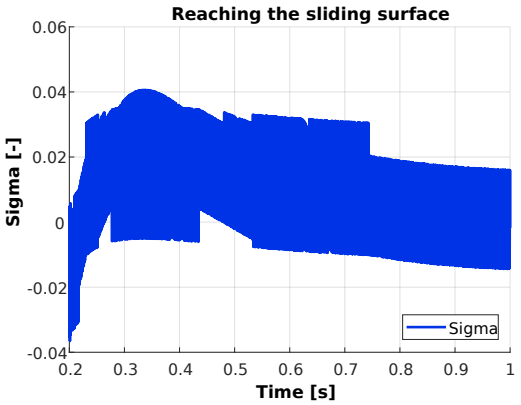


Figure 10: After reaching the sliding surface, the sliding mode becomes active.

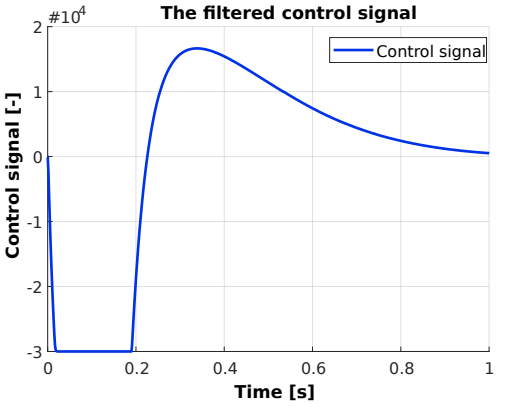


Figure 12: The filtered signal of the sliding-mode controller.

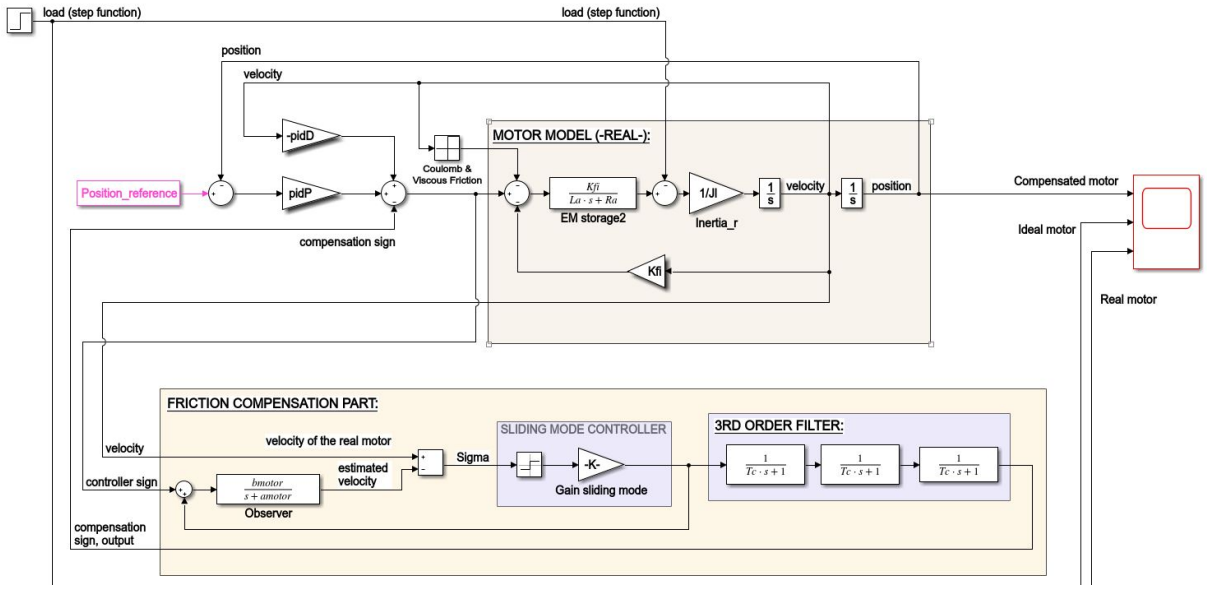


Figure 13: Full model of the compensated motor subjected to calibration.

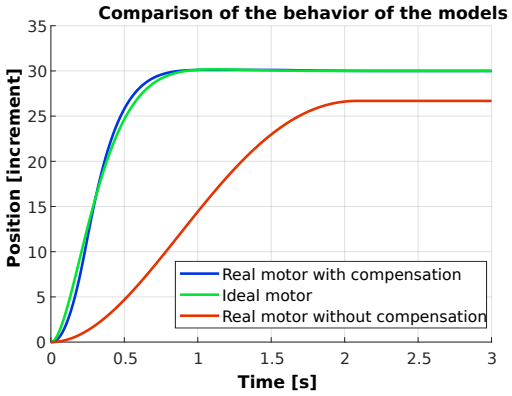


Figure 14: Position signals after Coulomb friction was accounted for in the system.

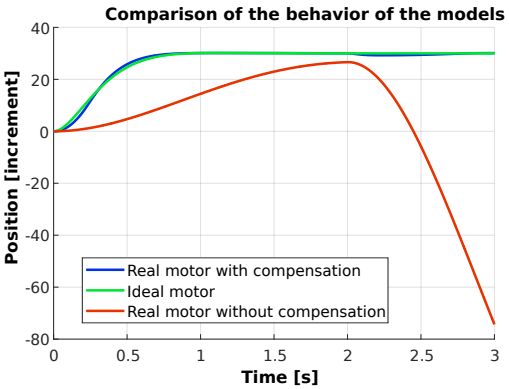


Figure 15: An additional load added to the system.

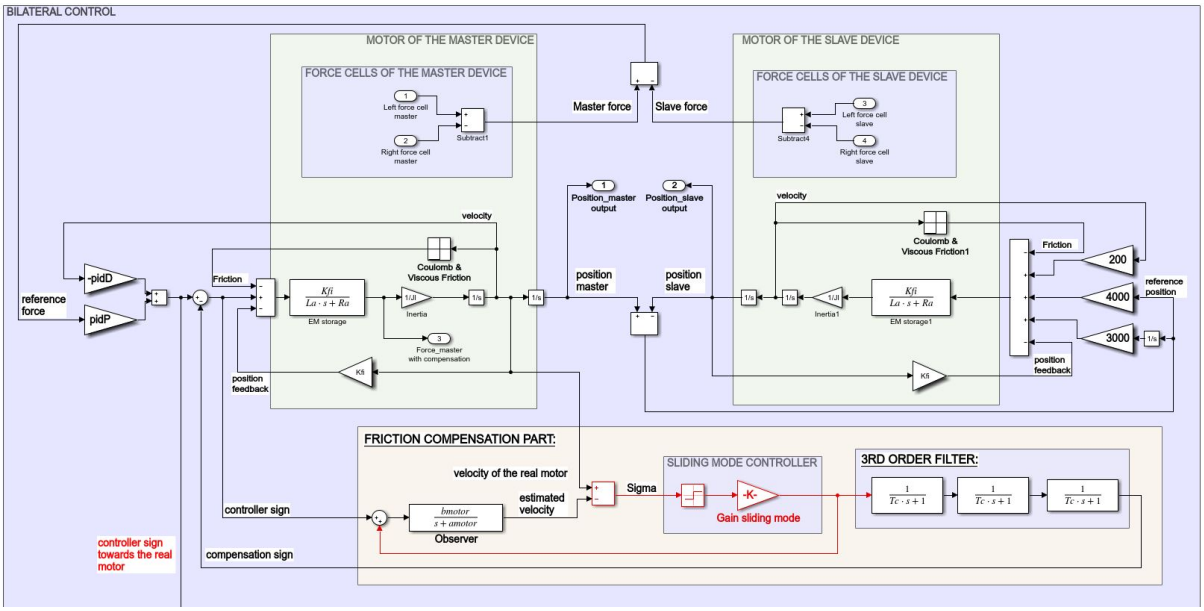


Figure 16: Bilateral control model.

Figure 17: The same parameters applied to the slave arm failed to yield results.

Figure 18: Position signal following of the bilateral system.

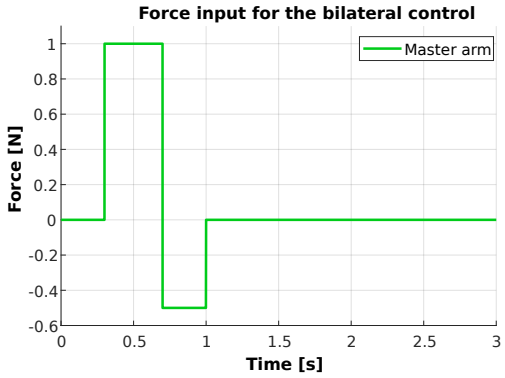


Figure 19: Force input to the bilateral control.

In Fig. 14 the position signals are shown after the addition of the Coulomb friction to the system. This new component has a stabilizing effect on the oscillation of the real motor which is greater than can be compensated for without the introduction of a permanent error. However, the behaviour of the motor compensated for by friction remains unchanged. The values of the Coulomb friction and viscous friction used in this experiment were $2,200$ and 0.3, respectively. Then after 2 seconds an additional load was added to the system (Fig. 15). It was noticeable that when the position signal of the normal motor started to rapidly decrease, the model compensated for by friction adapted to the new circumstances. The value of the load added after 2 s was 42.

4. Results and Discussion

4.1 Final model

Finally, bilateral haptic control was realized (Fig. 16). The model of the two force cells on the real-life device was added to the system by applying the same logic as used on the real. Each model of the motor was extended by a left- and right-force block.

In terms of bilateral control, the difference between the force blocks of each joystick was compared, which was also the input of the force controller of the master arm. Subsequently, the position of the master arm was compared to that of the slave arm, which was the input of the position controller of the slave arm (Fig. 16). It would be illogical to apply the same parameters to the position controller as to the one that accounts for compensation because without the component of compensation the friction force would prevent the proper position from being reached (Fig. 17).

Nevertheless, if an individually tuned proportional–integral–derivative (PID) controller is connected to the component responsible for position control of the system, the signal that results would be visible (Fig. 18). In this case, the input signal was applied to the right force cell of the master arm in a shape as is shown in Fig. 19.

It is impossible to eliminate the entire effects of friction and inertia on the mechanical construction, but they can be reduced significantly. The magnitude of the friction force exceeded the limit suitable for a smooth, comfortable operation. Compensation aims to make the system behave like an ideal model subject to a minimal level of friction. Due to the compensation, the operator can move the master joystick with ease.

Over a series of experiments, classical and sliding mode-based model reference adaptive control methods were compared. The applications of these methods with regard to compensation for friction are published in [11, 12].

It is crucial to identify an optimal adaptation parameter which facilitates rapid adaptation but avoids overcompensation. In the event of overcompensation, the joystick moves randomly even in the absence of reference torque because of the measurement noise. As the size of the adaptation parameter increases, adaptation occurs more rapidly, however, the likelihood of random movements rises.

5. Conclusion

An experimental telemanipulation system was presented in this paper. The master device was a serial linked lever-type haptic interface with force feedback. Even though it is impossible to eliminate the entire effects of friction and inertia on the mechanical construction, a reference model that accounts for minor levels of friction was designed. The system was forced to follow the reference model. In other words, the original dynamics of the master device were replaced by a virtual version which ensures a comfortable degree of manipulation for the operator.

REFERENCES

[1] Arai, F.; Sugiyama, T.; Fukuda, T.; Iwata, H.; Itoigawa K.: Micro tri-axial force sensor for 3D bio-micromanipulation, *1999 IEEE International Conference on Robotics and Automation*, 1999 **4**, pp. 2744–2749 ISBN: 0-7803-5180-0 DOI: 10.1109/ROBOT.1999.774012

[2] Kwon, D-S.; Woo, K. Y.; Cho, H. S.: Haptic control of the master hand controller for a microsurgical telerobot system, *1999 IEEE International Conference on Robotics and Automation*, 1999 **3**, pp. 1722–1727 ISBN: 0-7803-5180-0 DOI: 10.1109/ROBOT.1999.770357

[3] N. Ando, M. Ohta, H. Hashimoto: "Micro Teleoperation with Haptic Interface", *Proceedings of the 2000 IEEE Int. Conf. on Industrial Electronics, Control and Instrumentation*, pp. 13-18, 2000, 10.1109/IECON.2000.973119

[4] Yokokohji, Y.; Yoshikawa, T.: Bilateral control of master-slave manipulators for ideal kinesthetic coupling-formulation and experiment, *IEEE T. Robotics Autom.*, 1994 **10**(5), 849–858 DOI: 10.1109/70.326566

[5] Hashtrudi-Zaad, K.; Salcudean, S. E.: Analysis of control architectures for teleoperation systems with impedance/admittance master and slave manipulators, *Int. J. Robotics Res.*, 2001 **20**(6), 419–445 DOI: 10.1177/02783640122067471

[6] Korondi, P.; Young, K. D.; Hashimoto, H.: Sliding mode based disturbance observer for motion control In: Proceedings of the 37th IEEE Conference on Decision and Control, (IEEE Press, New York, USA) 1998, pp. 1926–1927 ISBN: 0-7803-4394-8 DOI: 10.1109/CDC.1998.758595

[7] V. I. Utkin: Sliding Modes in Control and Optimization, *Springer –Verlag*, 1992, ISBN: 978-3-642-84379-2 DOI: 10.1007/978-3-642-84379-2

[8] Ando, N.; Korondi, P.; Hashimoto, H.: Development of micromanipulator and haptic interface for networked micromanipulation, *IEEE/ASME T. Mechat.*, 2001 **6**(4), 417–427 DOI: 10.1109/3516.974855

[9] Korondi, P.; Young, K-K. D.; Hashimoto, H.: Discrete-time sliding mode based feedback compensation for motion control, *1996 IEEE International Workshop on Variable Structure Systems*, 1996, pp. 237–242 ISBN: 0-7803-3718-2 DOI: 10.1109/VSS.1996.578577

[10] Lukács P.: 1 degree of freedom haptic device for telemanipulation, *Master's Thesis*, 2015

[11] Ando, N.; Ohta, M.; Hashimoto, H.: Micro teleoperation with parallel manipulator, *2000 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2000 **1**, pp. 677–682 ISBN: 0-7803-6348-5 DOI: 10.1109/IROS.2000.894682

[12] Korondi, P.; Szemes, P. T.; Hashimoto, H.: Sliding mode friction compensation for a 20 DOF sensor glove, *J. Dyn. Sys., Meas., Control* , 2000 **122**(4) 611–615 DOI: 10.1115/1.1317232

HUNGARIAN JOURNAL OF
INDUSTRY AND CHEMISTRY
Vol. 47(1) pp. 49–56 (2019)
hjic.mk.uni-pannon.hu
DOI: 10.33927/hjic-2019-08

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

SIMULATION OF A BALANCED LOW-VOLTAGE ELECTRICAL GRID USING A SIMPLIFIED NETWORK MODEL

MÁRTON GREBER *[1] AND ATTILA FODOR[1]

[1]Department of Electrical Engineering and Information Systems, Faculty of Information Technology, University of Pannonia, Egyetem u. 10., Veszprém, H-8200, HUNGARY

A simulation method for low-voltage balanced distribution networks is proposed in this article. The novel method of node powers is based on the general calculation technique of node voltages. By researching only balanced networks, single-phase equivalents of the three-phase system are applicable. For the description of power lines, various parameters and matrices are available. In this work a simplified model is applied by using a purely resistive one. The active power results are solved through an iterative process. A main accomplishment is that the number of iterations needed is independent of the size of the network, and the process rapidly converges. Validation of the method is performed on the IEEE European Low-Voltage Test Feeder network. The simulation results confirm the achievements described in this paper.

Keywords: low-voltage distribution system, grid simulation, smart grid, method of node voltages, IEEE European Low-Voltage Test Feeder

1. Introduction

In recent years a lot of research has been conducted and progress made in the field of smart grid applications. Therefore, the demand for cloud-based systems with integrated simulation capabilities has increased. The calculation of the voltages, currents and powers of the components of the electrical (smart) grid is not easily achieved. This has resulted in the development of custom calculation methods which have the benefit of being fine-tuned for a particular application.

One of the fundamental network calculation methods - in a general sense - is the node voltages method. From a mathematical perspective, this method is based on solving a system of linear equations. As a result, this method can be implemented through various frameworks. One proposed solution revolves around using an open source discrete event simulator called OMNeT++ [1]. The models for electrical components need to be constructeda and the simulation is conducted via message handling. Another approach used Coloured Petri nets to model electrical networks [2]. A network model needs to be constructed for the simulation, which consists of a propagation process. A solution is presented for basic network types, while complex ones are calculated through decomposition. Both of these methods offer solutions but the method of node voltages regards currents as an input.

The more common approach to calculations of distribution systems calculation is via the method of power flow. This uses complex numbers to distinguish between active and reactive power. For the given nodes, both active and reactive power, voltage and phase angle are required to formulate the solution [3]. This poses a nonlinear problem, furthermore, the system of equations consists of real and imaginary subsets. Over the years, several pieces of research have dealt with this subject and the method known as DC power flow developed. By restricting the parameters the calculation was simplified, namely the voltage angles as well as generated and consumed active powers. This method assumes small differences in voltage angle and lossless lines [4]. The biggest downside of the method is that it cannot be used to calculate line losses because of the assumptions, although efforts are being made to overcome this obstacle [5].

By taking into account the extent of a given power system, different parameters of the power line model become dominant [6]. In low-voltage systems the distance between adjacent nodes is smaller than in high-voltage systems. By using this information a new set of restrictions is proposed, inspired by the DC power flow. It can be regarded as a complementary method since the line reactances are neglected. Utilising a resistive transmission line model offers the possibility to calculate distribution losses. This method has been developed from the method of node voltages.

In this article the issues of harmonic currents [7] and unbalanced networks [8] are not examined, it is assumed that the currents as well as voltages are sinusoidal and the grid is balanced with balanced three-phase loads. The

*Correspondence: greber.marton@virt.uni-pannon.hu

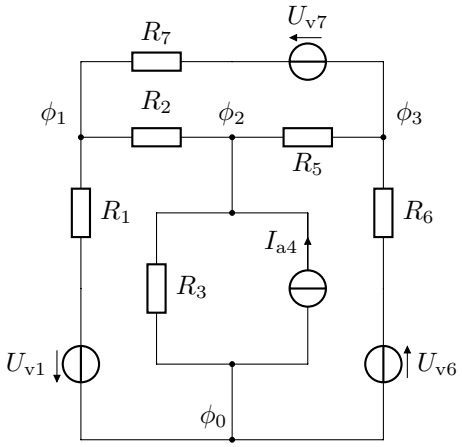


Figure 1: Example network

neutral wire can be omitted by restricting the set of networks in balanced low-voltage systems. This enables the application of a one-line equivalent circuit.

2. Method of node voltages

In distribution network calculations the main emphasis is on power flow. That means the voltages and angles are calculable when the generated and consumed powers are given [9]. The known formula for calculating the active power in a DC network is the following:

$$P = UI, \tag{1}$$

but the observed systems operate on alternating currents, hence the notion of AC power needs to be introduced [10]. Since voltages and currents are time-varying quantities, these are expressed as complex numbers, namely $\overline{U}$ and $\overline{I}$. The complex power is calculated by multiplying the conjugate of $\overline{I}$ by the voltage

$$\overline{S} = \overline{U}\,\overline{I}^{*}. \tag{2}$$

The phase angle of the complex power is defined by the difference between the angles of voltage and current:

$$\varphi = \arg(\overline{U}) - \arg(\overline{I}). \tag{3}$$

By observing the real and imaginary parts of the complex number, the active (P) and reactive powers (Q) are obtained:

$$P = \mathrm{Re}(\overline{S}),\ Q = \mathrm{Im}(\overline{S}). \tag{4}$$

A known method for the analysis of electrical networks is the method of node voltages. The component values, e.g. the sources of resistance, voltage and current are given, the unknown variables are the node voltages. The calculation steps are best explained through an example network as shown in Fig. 1.

One key component of the process is the definition of a directed graph for the example network shown in Fig. 2. The reference directions in the graph are arbitrary.

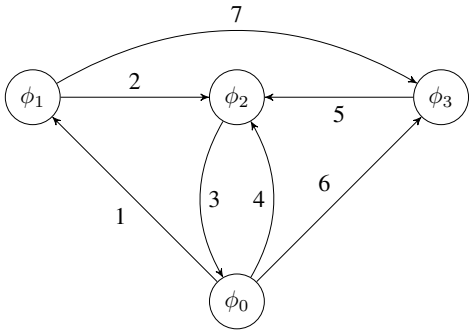


Figure 2: The reference directed graph

By using the graph in addition to Kirchoff's first law, the nodal equations can be written in the following form:

$$\begin{cases} I_1 - I_2 - I_7 = 0 \\ I_2 - I_3 + I_4 + I_5 = 0 \\ I_7 - I_5 + I_6 = 0 \end{cases} \tag{5}$$

After that, Ohm's law is used to express the edge currents with regard to the nodal voltages:

$$\begin{cases} I_1 = -G_1(\phi_1 - U_{v1}) \\ I_2 = G_2(\phi_1 - \phi_2) \\ I_3 = G_3\phi_2 \\ I_4 = I_{a4} \\ I_5 = G_5(\phi_3 - \phi_2) \\ I_6 = -G_6(\phi_3 + U_{v6}) \\ I_7 = G_7(\phi_1 + U_{v7} - \phi_3) \end{cases} \tag{6}$$

where the conductance of the arm of the network is denoted by G. If these steps are followed, the system of equations must be rearranged. The final form can be obtained by substituting the currents into Eq. (5) and rearranging it to determine the nodal voltages [11]:

$$\boldsymbol{\phi} = \boldsymbol{G}_{\mathbf{e}}^{-1}\boldsymbol{I}_{\mathbf{e}} \tag{7}$$

The notation represents the dimensions of the elements it contains: the node voltage vector is obviously denoted by $\boldsymbol{\phi}$, the nodal admittance matrix by $\boldsymbol{G}_{\mathbf{e}}$ and the excitation vector by $\boldsymbol{I}_{\mathbf{e}}$. At first sight, this equation does not account for the voltage sources in the branches. Since these elements can be described in terms of current dimensions, the excitation vector takes the following form:

$$\boldsymbol{I}_{\mathbf{e}} = \begin{bmatrix} -G_1U_{v1} + G_7U_{v7} \\ -I_{a4} \\ -G_7U_{v7} + G_6U_{v6} \end{bmatrix}. \tag{8}$$

2.1 The generalized nodal equations

The above-mentioned method describes the working principle of the technique, but it is not suitable for algorithmic applications. A generalized approach is needed. One has to define column vectors for the voltage sources:

$$\boldsymbol{U}_{\mathbf{v}}^{\top} = \begin{bmatrix} u_{v1} & u_{v2} & \dots & u_{vn} \end{bmatrix}, \tag{9}$$

current sources:

$$\boldsymbol{I}_{\mathbf{a}}^{\top} = \begin{bmatrix} i_{\mathrm{a}1} & i_{\mathrm{a}2} & \dots & i_{\mathrm{a}n} \end{bmatrix}, \tag{10}$$

and admittances:

$$\boldsymbol{Y}^{\top} = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix}, \tag{11}$$

where $\boldsymbol{Y}$ refers to the complex conductance known as the admittance. Each scalar of the vectors contains the aforementioned properties of a particular edge. For the representation of the graph, an incidence matrix ($\boldsymbol{A}$) is used of $(m-1) \times n$ dimensions, and describes connections between nodes and edges:

$$a_{i,j} = \begin{cases} 0, & \text{if } j \text{ is not connected to } i \\ 1, & \text{if } j \text{ is pointing away from } i \\ -1, & \text{if } j \text{ is pointing towards } i \end{cases}, \tag{12}$$

where $i = 1, 2, \dots, m-1$ represents the nodes of the graph and $j = 1, 2, \dots, n$ denotes the edge index set. As with the direction of the graph, the reference of $\boldsymbol{A}$ is also arbitrary, it is only manifested when multiplied by (-1). These matrices can be constructed algorithmically and facilitate the use of the following equation [12]:

$$\boldsymbol{\phi} = \boldsymbol{Y}_{\boldsymbol{A}}^{-1} \boldsymbol{A} \big(\mathrm{diag}(\boldsymbol{Y})\, \boldsymbol{U}_{\mathbf{v}} - \boldsymbol{I}_{\mathbf{a}} \big), \tag{13}$$

where:

$$\boldsymbol{Y}_{\boldsymbol{A}} = \boldsymbol{A}\, \mathrm{diag}(\boldsymbol{Y}) \boldsymbol{A}^{\top}. \tag{14}$$

A method which can be used to compute node voltages by defining a graph represented by the incidence matrix that includes the electrical properties of the edges is proposed as a result.

3. Network model

In order to reduce the amount of computational power that is needed for simulations, a simplified network model is used. The feeder points of low-voltage grids are transformer substations which convert forward the desired power from the medium voltage side into the base voltage, therefore, can be represented as voltage sources. Power flows from the feeder points to the customers via transmission lines. Compared to medium- or high-voltage lines, the length between consecutive nodes is smaller. As a result, these can be modelled as series resistances. Customers are represented as current sources with consumer references of course.

The computationally demanding part of the method of node voltages is to invert $\boldsymbol{Y}_{\boldsymbol{A}}$. In the case of complex analyses, the system of equations is separated into real and imaginary parts. Through one iteration cycle, two inverse calculations are needed. If voltages and currents are calculated as root mean square (RMS) values, the simulation method only requires real numbers. Therefore, the effort and time to calculate one cycle is halved.

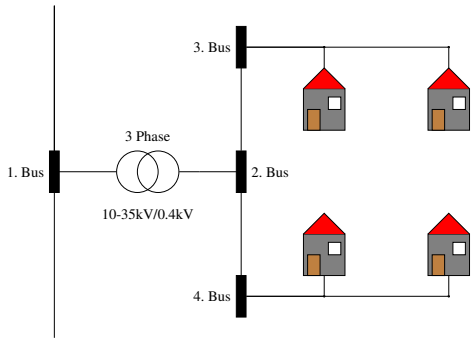


Figure 3: Low-voltage distribution system

3.1 Topology verification

In the structure of a low-voltage distribution network, some rules are noticeable. The transmission line forms a power rail to which the consumers can connect, as is shown in Fig. 3. These mainly consist of households with single-phase connections, i.e. one phase and the neutral wire are used [13].

The method of node voltages can be applied to general circuits, on the other hand, the proposed method of node powers can be applied to distribution networks. These form a tighter set, therefore, the topology needs to be checked to ensure it works properly.

If a network contains m nodes, m node equations can be obtained. On the other hand, only $m-1$ equations are linearly independent. Therefore, one node can be omitted, namely the 0 V node is omitted in the proposed methods. If an edge is connected to this point, it will have a non-zero column sum. If it is not connected to this point, it will have a column sum of zero. Using this, the criteria for the validation of topology can be formulated. The column sum for an arbitrary edge containing a current source in $\boldsymbol{A}$ cannot be equal to zero:

$$\forall i_{\mathrm{a}j} \neq 0, j \in \{1, \dots, n\} \rightarrow \sum_{i=1}^{m-1} a_{i,j} \neq 0. \tag{15}$$

Similarly, the column sum for an arbitrary edge containing a voltage source in $\boldsymbol{A}$ cannot be equal to zero:

$$\forall u_{vj} \neq 0, j \in \{1, \dots, n\} \rightarrow \sum_{i=1}^{m-1} a_{i,j} \neq 0. \tag{16}$$

In the case of an edge that possesses admittance, the column sum in $\boldsymbol{A}$ must be zero:

$$\forall y_j \neq 0, j \in \{1, \dots, n\} \rightarrow \sum_{i=1}^{m-1} a_{i,j} = 0. \tag{17}$$

In Algorithm 1, a pseudo code is shown that implements the above-mentioned criteria. It returns a Boolean value and is only true if the network topology is appropriate.

Algorithm 1 Topology verification

INPUT: $\boldsymbol{A}$, $\boldsymbol{Y}$, $\boldsymbol{I}_\mathbf{a}$, $\boldsymbol{U}_\mathbf{v}$
OUTPUT: *IsValidTopology*

IsValidTopology$= 1$
$n =$ NumberOfColumns($\boldsymbol{A}$)
for $i = 1$ to n **do**
  $s = 0$
  if $\boldsymbol{I}_\mathbf{a}(i)$!$= 0$ **then**
    $s = \mathrm{sum}(\boldsymbol{A}(:, i))$
    if $s == 0$ **then**
      IsValidTopology$= 0$
    end if
  end if
  if $\boldsymbol{U}_\mathbf{v}(i)$!$= 0$ **then**
    $s = \mathrm{sum}(\boldsymbol{A}(:, i))$
    if $s == 0$ **then**
      IsValidTopology$= 0$
    end if
  end if
  if $\boldsymbol{Y}(i)$!$= 0$ **then**
    $s = \mathrm{sum}(\boldsymbol{A}(:, i))$
    if s !$= 0$ **then**
      IsValidTopology$= 0$
    end if
  end if
end for

4. Method of node powers

Since the simulated households are single-phase consumers, the corresponding active power can be calculated as follows:

$$P = U\ I\ \cos(\varphi). \tag{18}$$

Let us define the power factor ($\boldsymbol{PF}$) and current-source active power ($\boldsymbol{P}_\mathbf{a}$) vectors for further matrix calculations. Single elements in both of these describe properties with regard to a single edge.

$$\boldsymbol{PF}^\top = \begin{bmatrix}\cos(\varphi_1) & \cos(\varphi_2) & \dots & \cos(\varphi_n)\end{bmatrix}, \tag{19}$$

$$\boldsymbol{P}_\mathbf{a}^\top = \begin{bmatrix}p_{\mathrm{a}1} & p_{\mathrm{a}2} & \dots & p_{\mathrm{a}n}\end{bmatrix}. \tag{20}$$

Voltage can be expressed as the electric potential difference between two points. If the incidence matrix represents node-edge relations, transposing it will describe edge-node relations. Multiplying it with the node voltages column vector will result in edge voltages:

$$\boldsymbol{U} = \boldsymbol{A}^\top \boldsymbol{\phi}. \tag{21}$$

By determining these notations and definitions, the basic equation of node voltage can be extended:

$$\begin{cases} \boldsymbol{\phi} = \boldsymbol{Y}_{\boldsymbol{A}}^{-1} \boldsymbol{A}\big(\mathrm{diag}(\boldsymbol{Y})\, \boldsymbol{U}_\mathbf{v} - \boldsymbol{I}_\mathbf{a}\big) \\ \boldsymbol{P}_\mathbf{a} = \mathrm{diag}(\boldsymbol{PF})\mathrm{diag}(\boldsymbol{A}^\top \boldsymbol{\phi})\boldsymbol{I}_\mathbf{a} \end{cases}, \tag{22}$$

In the case of the method of node voltages, the node voltages are calculated using the given current consumption. However, in the case of the method of node powers, the node voltages must be calculated with regard to the consumption, given in terms of the active power. Eq. 22 opens up the possibility of finding a solution following a trial and error procedure.

In other words, if only one current source in this system of equations is changed, all the node voltages will also be changed. Let us consider the case where the current of one particular customer is changed until its active power becomes equal to its actual value. Nonetheless, if a second consumer is to be set in the same fashion, the first one will be ruined. Through positive changes in current to the second source, a net drop in voltage will occur. Since power is the product of current and voltage, the current was untouched so the active power will be less. Therefore, it is also clear that if the second current source is decreased, the active power of the first source will exceed its actual value.

4.1 Constant iteration current

In the aforementioned problem, the solution is acquired through an iterative process to build up the unknown currents of the system gradually, instead of trying to determine them individually. It is necessary to choose a value of the current for the iterations: I_iter. In the first step, every element of the current vector is zero, however, if an edge contains a consumer, its value will be set as the iteration current. The problem, namely that by changing one current, all the node voltages will also change, still exists. However, if the iteration current is sufficiently small, the ability to build up the parameters from the ground is viable. To select a source for the actual iteration, a new variable was defined:

$$dPP = \frac{\boldsymbol{P}_\mathbf{a}^\mathrm{sim}}{\boldsymbol{P}_\mathbf{a}}. \tag{23}$$

This can be calculated for all consumer edges per iteration and provides information about how similar the simulated active power is to its desired value. It is obvious that the edge with the smallest dPP requires the highest degree of correction, therefore, it is incremented by I_iter. Consequently, in every iteration the branch with the minimum dPP needs to be identified, using a simple minimum search. After appropriate incrementation, the node voltages must be calculated in order to determine the new values of power. This procedure can be observed in Algorithm 2.

In an ideal case, the algorithm converges into the desired power vector and the following expression will be true:

$$\forall i \in \{1, \dots, k\} \rightarrow dPP_i = 1, \tag{24}$$

where k denotes the number of consumers in the grid. It is clear that the rate of convergence and the accuracy of the algorithm are heavily influenced by I_iter. If this rate is particularly small, the results will be precise ($dPP = 1$). However, in this case the number of iteration cycles will be enormous because the function described is analogous

Algorithm 2 Constant iteration current

INPUT: $\boldsymbol{A}$, $\boldsymbol{Y}$, $\boldsymbol{P_a}$, $\boldsymbol{U_v}$, $\boldsymbol{PF}$, I_{iter}, m, n
OUTPUT: $\boldsymbol{\phi}$, $\boldsymbol{I_a}$

$\boldsymbol{I_a} = \text{zeros}(n, 1)$
for $i = 1$ to n **do**
  if $\boldsymbol{P_a}(i)$!= 0 **then**
    $\boldsymbol{I_a}(i) = I_{\text{iter}}$
  end if
end for
$\boldsymbol{\phi} = \text{CalculateNodeVoltages}(\boldsymbol{A}, \boldsymbol{Y}, \boldsymbol{U_v}, \boldsymbol{I_a})$
$\boldsymbol{P_a^{\text{sim}}} = \text{diag}(\text{diag}(\boldsymbol{A}^\top \cdot \boldsymbol{\phi}) \cdot \boldsymbol{I_a}) \cdot \boldsymbol{PF}$
$[dPP, index] = \min(\boldsymbol{P_a^{\text{sim}}} / \boldsymbol{P_a})$
while $dPP < \epsilon$ **do**
  $\boldsymbol{I_a}(index) = \boldsymbol{I_a}(index) + I_{\text{iter}}$
  $\boldsymbol{\phi} = \text{CalculateNodeVoltages}(\boldsymbol{A}, \boldsymbol{Y}, \boldsymbol{U_v}, \boldsymbol{I_a})$
  $\boldsymbol{P_a^{\text{sim}}} = \text{diag}(\text{diag}(\boldsymbol{A}^\top \cdot \boldsymbol{\phi}) \cdot \boldsymbol{I_a}) \cdot \boldsymbol{PF}$
  $[dPP, index] = \min(\boldsymbol{P_a^{\text{sim}}} / \boldsymbol{P_a})$
end while

to $\frac{1}{x}$. This can be observed by determining the actual number of iterations needed for the process:

$$\sum_{i=1}^{k} \frac{I_{ai}}{I_{\text{iter}}} \tag{25}$$

as the desired current will consist of portions of I_{iter} on every edge. According to how the desired current is divided by the iteration current, overshoots are possible. Therefore, a value of ϵ is needed in order to secure a suitable exit condition for the loop.

Basically the blue plot represents a function similar to $f(x) = c \bmod x$, where c denotes a given number. In the case of the method described, c stands for a desired load current and x represents possible iteration currents used in the algorithm. According to how the desired current is divided by the actual iteration current, false values can be calculated. The periodic increase in the error is a property of the modulo operation, since the remainder increases until an integer multiple of x is identified. Then the error is equal to zero but begins to increase again. The red line represents the number of iterations needed to converge into a final solution. If $x \ll c$, the maximum "overshoot" by the modulo operator is relatively small, on the other hand, if $x < c$, more significant errors can occur. It is clear that in the first case many more iterations are necessary than in the second. Since c cannot be determined beforehand, unnecessarily large errors can occur which represents the weakness of the algorithm.

It is clear that a compromise must be made between the run-time and precision, as illustrated in Fig. 4.

4.2 Dynamic iteration current

To fix the weaknesses of the algorithm elaborated on in the previous chapter, an advanced version was developed. The first aspect, in which there is room for improvement, is to obtain reasonable initial values of $\boldsymbol{I_a}$. The process

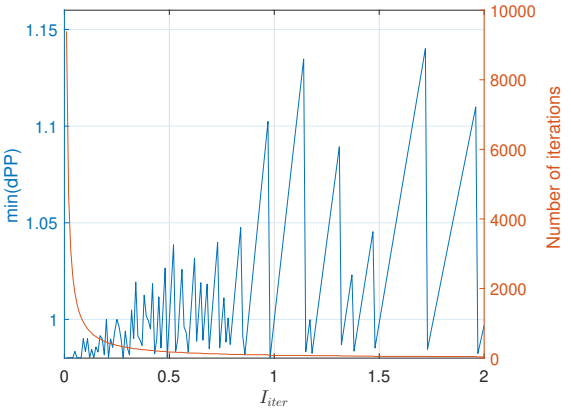


Figure 4: Minimum error - number of iterations

does not have to originate from $I_{ai} = 0$, consequently, a considerable amount of cycles can be skipped. In distribution networks the deviation from the nominal voltage (U_{n}) is always regulated by standards, for example, in Hungary it is approximately $\pm 7.5\%$. Using this restriction, a general estimation can be made for the nodes. The following $\min()$ and $\max()$ operators relate to the values of the given function. The possible interval between current values can be formulated as follows:

$$\begin{aligned} \min(\boldsymbol{I_a}) &= \frac{p_{\text{a}i}}{U_n \cos(\varphi_i)} \\ \max(\boldsymbol{I_a}) &= \frac{p_{\text{a}i}}{(1-D) U_{\text{n}} \cos(\varphi_i)} \end{aligned}, \tag{26}$$

where D is the aforementioned deviation value. This means overestimating the voltage results in the minimum value of the current. Setting the starting values of the currents according to the minimum approximation is adequate. To calculate the maximal remaining error, the minimum value of dPP_i needs to be calculated:

$$dPP_i = \frac{U_{\text{n}}(1-D)\frac{p_{\text{a}i}}{U_{\text{n}} \cos(\varphi_i)} \cos(\varphi_i)}{p_{\text{a}i}} = 1 - D. \tag{27}$$

The logic behind this equation is as follows: the fraction in the numerator is the minimum current estimation the $U_{\text{n}}(1-D)$, on the other hand, is the worst case scenario in terms of the voltage. The deviation is defined by D, so under no circumstances can the voltage drop below $U_{\text{n}}(1-D)$, once the initial value has been set. Because D is small, this approach alone solves a huge part of the problem, since by taking the Hungarian voltage levels as an example, a dPP value greater or equal to 0.925 is achieved!

Since the initial value problem has been solved, the remaining iterations can also be improved by taking the aforementioned method one step further. In order to take the absolute error in the active power into account, the following variable is introduced:

$$dP = P_{\text{a}} - P_{\text{a}}^{\text{sim}}. \tag{28}$$

which can also be formulated using the relative error:

$$dP = P_{\mathrm{a}}(1 - dPP). \tag{29}$$

Since the whole algorithm is founded on an iterative method, it would be suitable to use the minimum estimation process of the worst-case scenario in the following iterations. The idea in theory, however, is similar to the power value that is being approximated changes. Only the remaining error component needs to be recalculated. Therefore, the approximation will take the given dP_i into consideration instead of the whole $p_{\mathrm{a}i}$. U_{n} can still be used for this, but is not ideal. Once the initial values are set, a node voltage calculation is performed, thus the new node voltages can be used. As a result, the new values must be used as the upper limit of the voltage. Finally, the current generated for a particular consumer can be represented by the following series:

$$I_{\mathrm{a}i} = \frac{1}{\cos(\varphi_i)} \Bigg(\underbrace{\frac{p_{\mathrm{a}i}}{U_{\mathrm{n}}}}_{j=1} + \underbrace{\frac{dP_i'}{\phi_i'}}_{j=2} + \underbrace{\frac{dP_i''}{\phi_i''}}_{j=3} + \dots \Bigg), \tag{30}$$

where $j = 1, 2, \dots$ denotes the number of iterations. If $j \to \infty$, the simulated values approach the desired active powers. In theory this would mean that an infinite number of iterations would be necessary. The maximum error or minimum dPP can be calculated as shown before not only for the initial values but also for the upcoming iteration currents. Since in every consequent cycle the absolute error from the previous cycle is corrected, the remaining dP will be corrected by a minimum dPP equal to $(1-D)$. Since the fitted values are multiplied, the minimum value of the j-th iteration can be calculated as follows:

$$dPP_{\min} = 1 - D^j. \tag{31}$$

The process of the method of dynamic node powers is shown in Algorithm 3. In the literature review, the

Algorithm 3 Dynamic iteration current

INPUT: U_{n}, $\boldsymbol{A}$, $\boldsymbol{Y}$, $\boldsymbol{P_{\mathrm{a}}}$, $\boldsymbol{U_{\mathrm{v}}}$, $\boldsymbol{PF}$
OUTPUT: $\boldsymbol{\phi}$, $\boldsymbol{I_{\mathrm{a}}}$

$\boldsymbol{\phi} = \mathrm{ones}(m, 1) \cdot U_{\mathrm{n}}$
$\boldsymbol{I_{\mathrm{a}}} = \boldsymbol{P_{\mathrm{a}}} / \left((\boldsymbol{A}^{\top} \cdot \boldsymbol{\phi}) \cdot \boldsymbol{PF} \right)$, if $\frac{0}{0} : \ \boldsymbol{I}_{\mathrm{a}i} \to 0$
$\boldsymbol{\phi} = \mathrm{CalculateNodeVoltages}(\boldsymbol{A}, \boldsymbol{Y}, \boldsymbol{U_{\mathrm{v}}}, \boldsymbol{I_{\mathrm{a}}})$
$\boldsymbol{P}_{\mathrm{a}}^{\mathrm{sim}} = \mathrm{diag}(\mathrm{diag}(\boldsymbol{A}^{\top} \cdot \boldsymbol{\phi}) \cdot \boldsymbol{I_{\mathrm{a}}}) \cdot \boldsymbol{PF}$
$dPP = \min(\boldsymbol{P}_{\mathrm{a}}^{\mathrm{sim}} / \boldsymbol{P_{\mathrm{a}}})$
while $dPP < \epsilon$ **do**
    $I_{\mathrm{iter}} = (\boldsymbol{P_{\mathrm{a}}} - \boldsymbol{P}_{\mathrm{a}}^{\mathrm{sim}}) / (\boldsymbol{U} \cdot \boldsymbol{PF})$, if $\frac{0}{0} : \ \boldsymbol{I}_{\mathrm{a}i} \to 0$
    $\boldsymbol{I_{\mathrm{a}}} = \boldsymbol{I_{\mathrm{a}}} + I_{\mathrm{iter}}$
    $\boldsymbol{\phi} = \mathrm{CalculateNodeVoltages}(\boldsymbol{A}, \boldsymbol{Y}, \boldsymbol{U_{\mathrm{v}}}, \boldsymbol{I_{\mathrm{a}}})$
    $\boldsymbol{P}_{\mathrm{a}}^{\mathrm{sim}} = \mathrm{diag}(\mathrm{diag}(\boldsymbol{A}^{\top} \cdot \boldsymbol{\phi}) \cdot \boldsymbol{I_{\mathrm{a}}}) \cdot \boldsymbol{PF}$
    $dPP = \min(\boldsymbol{P}_{\mathrm{a}}^{\mathrm{sim}} / \boldsymbol{P_{\mathrm{a}}})$
end while

Newton-Raphson method was examined in more detail which uses the mismatch in power by incorporating the Jacobian matrix in order to achieve convergence. The proposed error recalculation method was developed whilst taking that process into consideration. Another possible solution to the calculation of powers could be an algorithm, in which the total power is recalculated rather than the power mismatch. The explicit expression for convergence was determined first, and the total power recalculation method did not suggest better results.

Although through simulation and observation, it was noted that both require the same amount of iterations. Using the Monte Carlo method, a large number of random networks were created in order to monitor the run-time. Results showed the presence of slight deviations with regard to each other. Nevertheless, the two methods require almost the same amount of computational time, as is shown in Fig. 5.

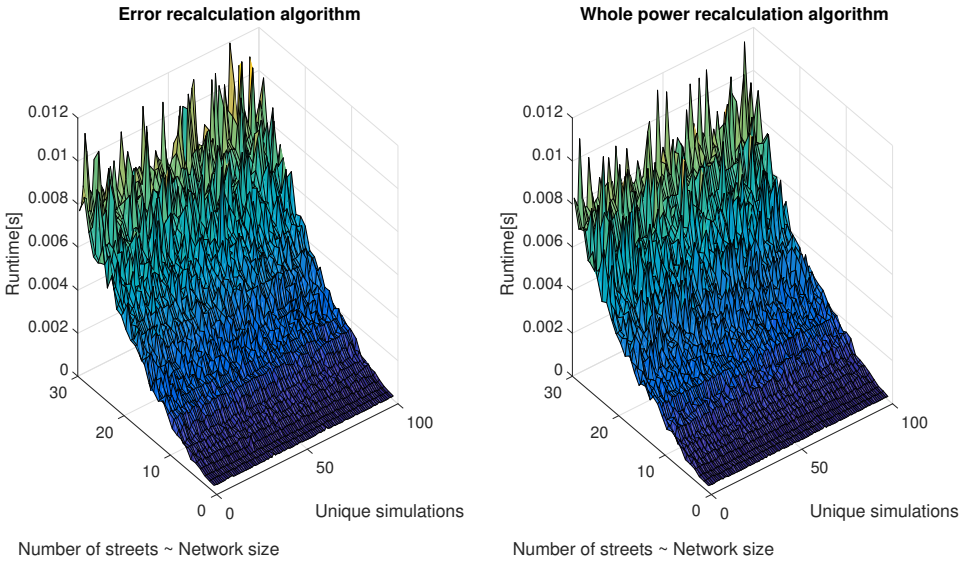


Figure 5: Monte Carlo simulation

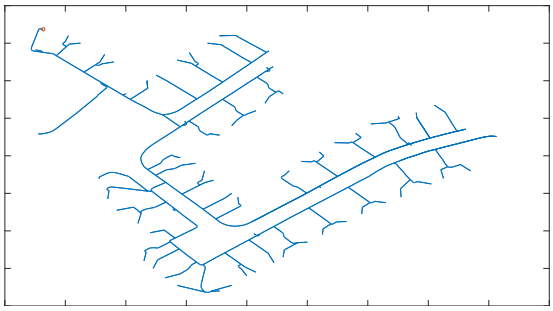


Figure 6: The topology of the test network

5. Simulation results

In order to verify the proposed method properly, a reference network with available simulation data was required. For verification purposes the IEEE European Low-Voltage Test Feeder network [14] was used. The simulation process consisted of the implementation of the stated algorithm and importation of network data using MATLAB. The topology of the test grid is shown in Fig. 6.

Some of the important network parameters are the following: it consists of 906 nodes which are connected by 927 edges that supply 55 single-phase consumers. Previous runs have shown that approximately 2 iterations are sufficient for engineering purposes. For the sake of accuracy, a limit was set for dPP, therefore ϵ is still shown in Algorithm 3.

The minimum estimations of dPP predicted that the error should rapidly converge to zero. The simulation verified this statement, the rapid decrease in the absolute error can be seen in Tables 1 - 3. It can be clearly observed that in the case of a network of such a size, the maximum voltage difference can be maintained under 10 mV (Table 4).

The results of the simulation can be seen in Fig. 7 which represents the bus voltages of a single phase. Note that the y axis is divided into increments of 50 mV.

Table 1: Convergence of dP in Phase 'A'

Iteration	$\max(dP)$
0.	0.052137599675 W
1.	0.000041085202 W
2.	0.000000032240 W

Table 2: Convergence of dP in Phase 'B'

Iteration	$\max(dP)$
0.	0.042119472811 W
1.	0.000029088116 W
2.	0.000000019550 W

Table 3: Convergence of dP in Phase 'C'

Iteration	$\max(dP)$
0.	0.043952897250 W
1.	0.000025174297 W
2.	0.000000014771 W

Table 4: Maximum voltage differences

Phase 'A'	Phase 'B'	Phase 'C'
0.056859 V	0.036058 V	0.023649 V

Since the active power values are reached with an excellent degree of precision, one would expect that the voltage differences would be smaller. However, this effect does not originate from the algorithm, rather from the simplified network model. This facilitates the possibility of achieving small simulation times. Since the introduction of the method of dynamic iteration currents, the number of iterations is independent of the size of the network. The computationally heavy component is the calculation of the inverse of $\boldsymbol{Y_A}$. The topology of the network remains unchanged during the iterations. Therefore, the steady state simulation of a three-phase network requires, independen of the network size, only three matrix inversions. The duration of the simulation of this network was 2.29 s at a particular instant.

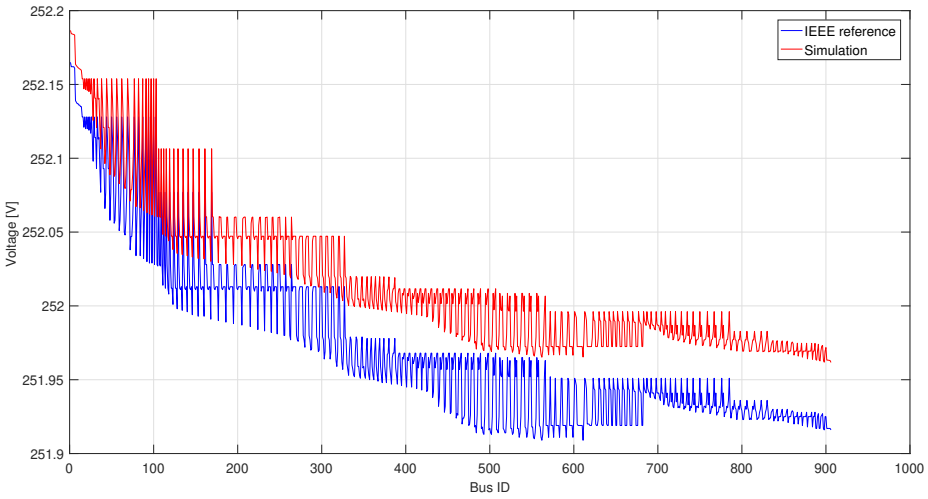


Figure 7: Simulation results for phase 'A'

6. Conclusion

A novel simulation method for low-voltage distribution networks is proposed in this paper. The existing method of node voltages is further developed in order to handle active power calculations in low-voltage grids. These systems consist of a unique topology to which the process was fitted. The verification criteria for the network structure was formulated. A solution was proposed in which the number of iterations is independent of the size of the network and the simulation error decreases exponentially. The method was tested and verified on the IEEE European Low-Voltage Test Feeder network. This confirmed the statements about the algorithm. Insignificant errors appeared in the test results but these were the effect of the simplified network model.

In the future, the algorithm could be improved to handle unbalanced distribution networks. With the aid of appropriate modifications, distributed generation could be taken into consideration that accounts for not only power consumption but also generation. The algorithm can serve as a foundation of network diagnostics by using it to detect faults as well as technical or non-technical losses.

Acknowledgement

We acknowledge the financial support of Széchenyi 2020 under the EFOP-3.6.1-16-2016-00015. We acknowledge the financial support of Széchenyi 2020 under the GINOP-2.2.1-15-2017-00038.

Notations

$\boldsymbol{\phi}$	Node-voltage vector
$\boldsymbol{I_e}$	Excitation vector
$\boldsymbol{U_v}$	Voltage-source vector
$\boldsymbol{I_a}$	Current-source vector
$\boldsymbol{A}$	Incidence matrix
$\boldsymbol{Y}$	Admittance matrix
$\boldsymbol{Y_A}$	Nodal admittance matrix
m	Number of nodes
n	Number of edges
$\cos(\varphi)$	Power factor
$\boldsymbol{PF}$	Power factor vector
$\boldsymbol{P_a}$	Current-source active-power vector
$\boldsymbol{U}$	Branch voltage vector
dPP	Delta power percentage
I_{iter}	Iteration current
dP	Delta power
D	Voltage level deviation
$\boldsymbol{P}_{\mathbf{a}}^{\text{sim}}$	Simulated power vector
S	Complex power
P	Active power
Q	Reactive power
φ	Phase angle
G	Conductance
ϵ	Threshold value for iteration

REFERENCES

[1] Sőrés, M.; Fodor, A.: Simulation of electrical grid with OMNeT++ open source discrete event system simulator, *Hung. J. Ind. Chem.*, 2016 **44**(2), 85–91, DOI: 10.1515/hjic-2016-0010

[2] Pózna, A.I.; Fodor, A.; Gerzson, M.; Hangos, K.M.: Colored Petri net model of electrical networks for diagnostic purposes, *IFAC-PapersOnLine*, 2018 **51**(2), 260–265, DOI: 10.1016/j.ifacol.2018.03.045

[3] Stagg, G.; El-Abiad, A.H.: Computer methods in power system analysis (McGraw-Hill), interntaional student edn., 1968

[4] Hertem, D.V.; Verboomen, J.; Purchala, K.; Belmans, R.; Kling, W.L.: Usefulness of DC power flow for active power flow analysis, *The 8th IEE International Conference on AC and DC Power Transmission*, 2006 **1**, DOI: 10.1049/cp:20060013

[5] Stott, B.; Jardim, J.; Alsac, O.: DC power flow revisited, *IEEE T. Power Syst.*, 2009 **24**(3), 1290–1300, DOI: 10.1109/TPWRS.2009.2021235

[6] Das, J.C.: Load flow optimazitaion and optimal power flow (CRC Press), 2017

[7] Görbe, P.; Magyar, A.; Hangos, K.M.: THD reduction with grid synchronized inverter's power injection of renewable sources, in 2010 International Symposium on Power Electronics Electrical Drives Automation and Motion (SPEEDAM) (IEEE), 1381–1386, DOI: 10.1109/SPEEDAM.2010.5545079

[8] Neukirchner, L.; Görbe, P.; Magyar, A.: Voltage unbalance reduction in the domestic distribution area using asymmetric inverters, *J. Clean. Prod.*, 2017 **142**, 1710–1720, DOI: 10.1109/59.575728

[9] Kersting, W.H.: Distribution system modeling and analysis (CRC Press), 2016, ISBN: 9781439856475

[10] Bird, J.: Electrical circuit theory and technology (Taylor & Francis), 4th edn., 2010, ISBN: 185617770X

[11] Wang, X.F.; Song, Y.; Irving, M.: Modern power system analysis (Springer), 2008, DOI: 10.1007/978-0-387-72853-7

[12] Zhang, F.; Cheng, C.S.: A modified Newton method for radial distribution system power flow analysis, *IEEE T. Power Syst.*, 1997 **12**(1), 389 – 397, DOI: 10.1109/59.575728

[13] Westinghouse Electric Corporation, Electrical transmission and distribution reference book (East Pittsburgh, PA), 1964

[14] Schneider, K.P.; Mather, B.A.; Pal, B.C.; Ten, C.W.; Shirek, G.J.; Zhu, H.; Fuller, J.C.; Pereira, J.L.R.; Ochoa, L.F.; de Araujo, L.R.; Dugan, R.C.; Matthias, S.; Paudyal, S.; McDermott, T.E.; Kersting, W.: Analytic considerations and design basis for the IEEE distribution test feeders, *IEEE T. Power Syst.*, 2017 **33**(3), 3181–3188, DOI: 10.1109/TPWRS.2017.2760011

HUNGARIAN JOURNAL OF
INDUSTRY AND CHEMISTRY
Vol. 47(1) pp. 57–63 (2019)
hjic.mk.uni-pannon.hu
DOI: 10.33927/hjic-2019-09

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

MODELING AND CALCULATION OF THE GLOBAL SOLAR IRRADIANCE ON SLOPES

ROLAND BÁLINT *[1], ATTILA FODOR[1], ISTVÁN SZALKAI[2], ZSÓFIA SZALKAI[3], AND ATTILA MAGYAR[1]

[1]Department of Electrical Engineering and Information Systems, Faculty of Information Technology, University of Pannonia, Egyetem u. 10, Veszprém, 8200, HUNGARY
[2]Department of Mathematics, Faculty of Information Technology, University of Pannonia, Egyetem u. 10, Veszprém, 8200, HUNGARY
[3]Eötvös Loránd University, Egyetem tér 1-3, Budapest, 1053, HUNGARY

The first step with regard to a simple model of a Photovoltaic Power Plant is developed in this paper based on astronomical and engineering principles. A solar irradiance model is presented in this paper that can be used to forecast the solar energy a surface on Earth is exposed to. The obtained model is verified against engineering expectations. The developed model can serve as a basis for forecasting the power of solar energy.

Keywords: solar irradiance, modeling, solar geometry, inclined surface

1. Introduction

As sustainable development is becoming highly important all over the world, the power production forecast of Photovoltaic Power Plants (PVPPs) is becoming ever more necessary. The Sun is the greatest potential energy source available. The thermal energy it emits can be transformed into electrical energy, or its irradiance can be converted into electricity by solar panels. The amount of energy produced by both of the solutions above is directly dependent on the solar irradiance.

Several works have focused on the determination or forecast of this energy and its dependence on the geometrical parameters of the surface. A general model of the angle of incidence of solar irradiance with regard to the azimuth is derived in Ref. [1] that is used for both fixed and tracking surfaces. In Ref. [2], a simple spectral model is given that is applicable to inclined surfaces and clear skies.

A very good comparative study was proposed in Refs. [3] and [4] where different area-specific models were analysed with regard to the optimal tilt angle. Unfortunately, the Hungarian region is not covered by the models used in this study. A semi-empirical method for short-term (hourly) solar irradiance forecasting was proposed in Ref. [5] founded on the widely used Ångström-Prescott model.

The aim of this work is to propose a model that is not only able to handle clear-sky conditions but clouds as well. The computational complexity of the model is also a key factor which should be kept to a minimum.

The organization of this paper is as follows: Section 2 introduces the problem at hand and also presents the motivation of the paper. Afterwards, the elementary astronomical relationships are collected and the proposed model is developed in Section 3. The mathematical model is verified against basic engineering expectations in Section 4. Finally, some concluding remarks are presented.

2. Problem statement

The produced energy forecast of the renewable energy sources is one of the main problems of the Distributed Energy Resource Management System. The operators of the electrical grids need precise mathematical models of renewable power plants, e.g. PVPPs, wind farms or hydroelectric power plants. The first step to develop a simple model of a PVPP is presented in this paper based on first astronomical then engineering principles.

The energy production of a PVPP is a function of the following:

- parameters of the PVPP,
- weather (cloud, temperature, wind),
- quality of the atmosphere (dust, relative humidity),
- activity of the Sun (11-year solar cycle).

The parameters of the PVPPs are the following:

- position,

*Correspondence: balazs.istvan@virt.uni-pannon.hu

- orientation of the PV solar panels,
  - tilt angle,
  - azimuth angle,
- surface area (or the number of PV solar panels),
- nominal power of the PV solar panels,
- efficiency of the solar inverter.

By using the first four parameters, the solar energy can be calculated by the developed solar irradiance model which is presented in this paper.

3. Astronomical relationships - Sun-Earth geometries

To estimate the solar irradiance from the Sun, some astronomical (geometrical) calculations to define the position of the Sun relative to the local horizontal surface need to be conducted. Many different equations are found in the literature to calculate the same parameter. The size of these equations, in addition to the frequency of measurements, vary significantly. In this section some equations are compared.

3.1 Comparison of the available solar geometry models

Due to the elliptic orbit of the Earth around the Sun, some parameters exist with a 1-year cycle. These parameters can be calculated with different equations and the difference between values of these equations varies due to discrepancies in the description of the method. The calculations regard a year as consisting of 365 days. Most of the results were obtained from Ref. [6] and references therein.

Relative Sun-Earth distance

The first parameter is the reciprocal of the square of the Sun-Earth distance during the year. The elliptic orbit of the Earth causes this to change in value. This relative distance can be calculated from equations

$$\begin{aligned} E_0 = \left(\frac{r_0}{r}\right)^2 = 1.000110 + \\ + 0.034221 \cos\left(2\pi\frac{d_n - 1}{365}\right) + \\ + 0.001280 \sin\left(2\pi\frac{d_n - 1}{365}\right) + \\ + 0.000719 \cos\left(4\pi\frac{d_n - 1}{365}\right) + \\ + 0.000077 \sin\left(4\pi\frac{d_n - 1}{365}\right) \end{aligned} \quad (1)$$

and

$$E_0 = \left(\frac{r_0}{r}\right)^2 = 1 + 0.033 \cos\left(2\pi\frac{d_n}{365}\right) \quad (2)$$

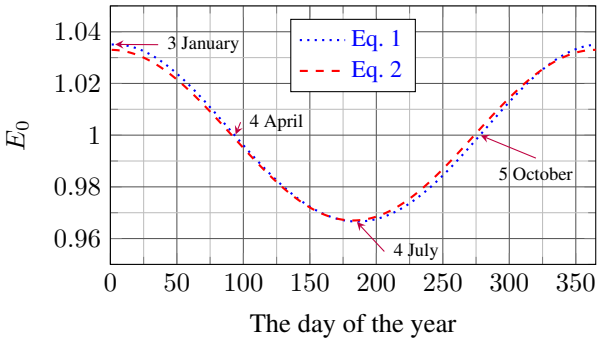


Figure 1: The calculated value of the reciprocal of the square of the Sun-Earth distance (E_0) using Eqs. 1 and 2

Eqs. 1 and 2 are from Refs. [6] and [7]. E_0 denotes the reciprocal of the square of the Sun-Earth distance, r_0 represents the mean Sun-Earth distance (1 AU) and r stands for the Sun-Earth distance on the d_nth day of the year. Values of E_0 using various equations are shown in Fig. 1.

This change will change the solar constant too. The solar constant, I_0, yields the following solar irradiance value (in W/m^2) at the edge of the Earth's atmosphere. This constant value is 1367 W/m^2 if the Earth is a distance of 1 AU (astronomical unit) from the Sun. This value is greater at perihelion (minimum Sun-Earth distance $\approx$ 0.983 AU on 3 January) and less at aphelion (maximum Sun-Earth distance $\approx$ 1.017 AU on 4 July). Since the irradiance is proportional to the surface area, which is inversely proportional to the square of the distance, the corrected solar irradiance can be calculated from

$$I_n = I_0 E_0. \quad (3)$$

Solar declination

The declination of the Sun, δ, yields the latitude above which the path of the Sun follows. Therefore, the zenith angle is zero at local noon. This value can be calculated from the following equations [6]:

$$\begin{aligned} \delta = \Bigg(0.006918 - \\ - 0.399912 \cos\left(2\pi\frac{d_n - 1}{365}\right) + \\ + 0.070257 \sin\left(2\pi\frac{d_n - 1}{365}\right) - \\ - 0.006758 \cos\left(4\pi\frac{d_n - 1}{365}\right) + \\ + 0.000907 \sin\left(4\pi\frac{d_n - 1}{365}\right) - \\ - 0.002697 \cos\left(6\pi\frac{d_n - 1}{365}\right) + \\ + 0.00148 \sin\left(6\pi\frac{d_n - 1}{365}\right) \Bigg) \frac{180^\circ}{\pi} \end{aligned} \quad (4)$$

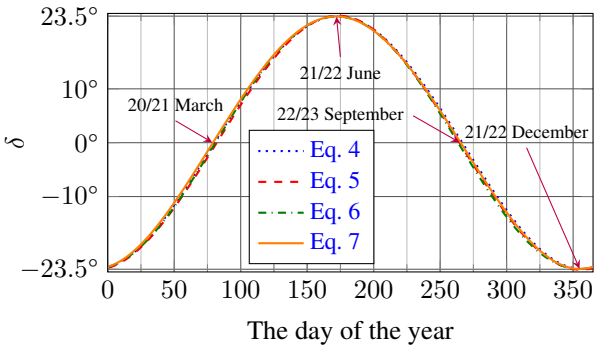


Figure 2: The calculated value of the declination of the Sun (δ) using Eqs. 4–7

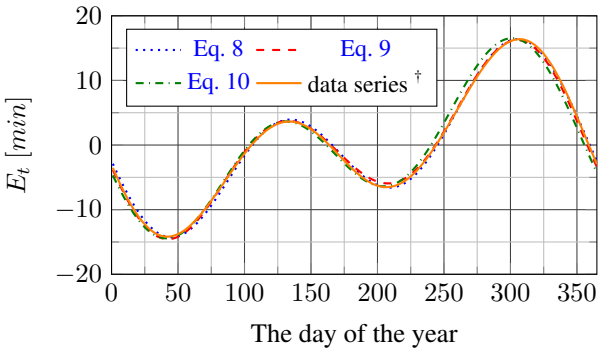


Figure 3: The values of the Equation of Time over a year using Eqs. 8–10 and a series for the year 2018

$$\delta = \arcsin\left(0.4\sin\left(\frac{360}{365}(d_n - 82)\right)\right) \tag{5}$$

$$\delta = 23.45° \sin\left(\frac{2\pi}{365}(d_n + 284)\right) \tag{6}$$

and

$$\delta = \arcsin\Big(0.39779\sin\big(4.8888 + 0.017214d_n + \\ + 0.033\sin(0.017214d_n)\big)\Big) \tag{7}$$

The results of different equations that calculate the declination of the Sun are shown in Fig. 2. The Sun is above the equator on 20/21 March (vernal equinox) and 22/23 September (autumnal equinox), moreover, the winter solstice is on 21/22 December and the summer solstice on 21/22 June. The difference between the equations in Fig. 2 is minimal. The maximum change in the declination when calculated on a daily basis is less than $0.5° \left(|\delta(d_{n+1}) - \delta(d_n)|\right)$.

Equation of time

A solar day varies in length throughout the year. A day on Earth is 24 hours long and is based on the rotation of the Earth around its polar axis. A solar day is based on the rotation of the Earth around its polar axis and on its motion around the Sun. The elliptic orbit of the Earth causes a difference between the time of day and the solar time of day (astronomical). This means that noon in local and astronomical times (the Sun is in the southern hemisphere) differ.

This time difference can be calculated from the following equations [6]:

$$E_t = \Big[0.000075 + 0.001868\cos\left(2\pi\frac{d_n - 1}{365}\right) - \\ - 0.032077\sin\left(2\pi\frac{d_n - 1}{365}\right) - \\ - 0.014615\cos\left(4\pi\frac{d_n - 1}{365}\right) - \\ -0.04089\sin\left(4\pi\frac{d_n - 1}{365}\right)\Big] 229.18 \tag{8}$$

$$E_t = \Big[0.043\sin\Big(2\big(4.8888 + 0.017214d_n + \\ + 0.033\sin(0.017214d_n)\big)\Big) - \\ - 0.03342\sin(0.017214d_n) + \\ + 0.2618t_{\mathrm{UTC}} - \pi\Big] 229.18 \tag{9}$$

and

$$E_t = -7.655\sin\left(2\pi\frac{d_n}{365}\right) + \\ + 9.873\sin\left(4\pi\frac{d_n}{365} + 3.588\right). \tag{10}$$

Eqs. 8 and 9 yield E_t in radians and the constant 229.18 converts this into minutes. This constant can be calculated from

$$229.18 = \frac{360°}{2\pi}\frac{360°}{24 \cdot 60\ \mathrm{min}}\ 16\ \mathrm{min} \tag{11}$$

Fig. 3 shows the results according to Eqs. 8–10 over a whole year, together with astronomical data for the year 2018.

The largest difference was calculated by Eq. 10 when compared with the other data.

3.2 Solar beam equations

Using the equations in Section 3.1, the position and angles of the Sun at a local (geological) position can be calculated with the GPS coordinates ϕ_{lat} and ϕ_{long}.

Fig. 4 shows the path of the Sun and the angles of the actual position of the Sun. The nomenclature is as follows:

- α: solar altitude angle
- θ_z: solar zenith angle
- ψ: solar azimuth angle

Since the altitude (α) and zenith angles (θ_z) are complementary angles,

$$\alpha + \theta_z = 90°, \tag{12}$$

the local solar hour angle can be calculated from [6]

†http://www.ppowers.com/EoT.htm

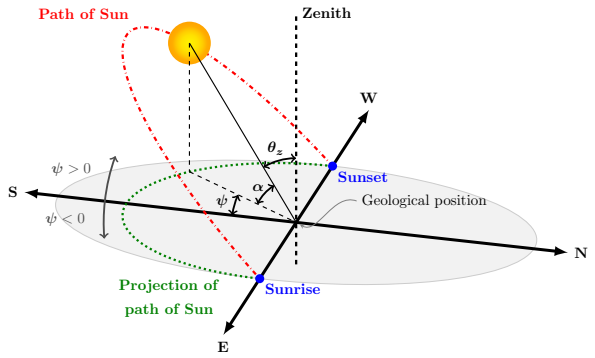


Figure 4: The local position of the Sun with important angles

$$h = 180° - 15° \left(t_{\mathrm{UTC}} + \frac{E_t}{60} \right) - \phi_{\mathrm{long}}, \tag{13}$$

where t_{UTC} is the local time in time zone $+0$ in hours and ϕ_{long} is the local longitudinal position in degrees.

The cosine of the solar zenith angle can be calculated by using the solar declination, local latitudinal position and solar hour angle. According to Eq. 12,

$$\cos(\theta_{\mathrm{z}}) = \sin(\delta)\sin(\phi_{\mathrm{lat}}) + \\ + \cos(\delta)\cos(\phi_{\mathrm{lat}})\cos(h), \tag{14}$$

and

$$\cos(\theta_{\mathrm{z}}) = \sin(\alpha), \tag{15}$$

(Eq. 14 is from Ref. [6]) the solar azimuth angle (ψ) can be calculated from

$$\cos(\psi) = \frac{\sin(\alpha)\sin(\phi_{\mathrm{lat}}) - \sin(\delta)}{\cos(\alpha)\cos(\phi_{lat})} \tag{16}$$

and

$$\sin(\psi) = \frac{\cos(\delta)\sin(h)}{\cos(\alpha)}. \tag{17}$$

3.3 Global solar irradiance

The solar irradiance can be calculated in a simple form:

$$I = I_n q^{T_{\mathrm{m}} z} \sin(\alpha), \tag{18}$$

where q denotes the clear sky transmissivity through a thickness of one atmosphere, T_{m} represents the Linke turbidity factor of the air and z stands for the relative path length of the solar beam through the atmosphere. Parameter q is a constant with a value of 0.93 (when the sky is dry and clear and the zenith angle is equal to zero). The value of T_{m} determines how many clear atmospheres should be stacked to achieve the actual spectral transmittance. This value depends on the amount of water vapor, dust, smog, etc. In the EU, T_{m} is between 1.8 and 4. The suggested formula for the calculation of parameter z is

$$z = \frac{l_{\mathrm{atm}}}{R_{\mathrm{Earth}} - r_{\mathrm{Earth}}}, \tag{19}$$

where l_{atm} denotes the path length of the solar beam through the atmosphere, r_{Earth} represents the mean radius of the Earth's surface and R_{Earth} stands for the radius of the Earth when the atmosphere is included ($R_{\mathrm{Earth}} - r_{\mathrm{Earth}}$ is the thickness of the atmosphere). The value of l_{atm} depends on the solar altitude and can be calculated using the law of cosines which yields

$$l_{\mathrm{atm}}^2 - 2\, r_{\mathrm{Earth}}\, l_{\mathrm{atm}} \cos(\alpha + 90°) + \\ + (r_{\mathrm{Earth}}^2 - R_{\mathrm{Earth}}^2) = 0. \tag{20}$$

The correct value of l_{atm} is the positive root of Eq. 20.

Eq. 18 yields the solar irradiance under ideal conditions. However, aerosols in the atmosphere cause diffuse scattering and absorb a proportion of the solar irradiance. Furthermore, clouds also have an effect on the irradiance. Some models divide the global solar irradiance into two parts:

- Direct solar irradiance
- Diffuse solar irradiance

Irradiance on a horizontal surface

Various models exist to calculate the global solar irradiance on a horizontal surface, e.g., a Hungarian model [8] is presented in

$$G = \Big(I_n \sin(\alpha) + a \Big) \Big(1 + b_1 N^{b_2} \Big), \tag{21}$$

where I_n denotes the corrected solar constant, N represents the cloudy parameter between 0 and 1 (0 when clear, 1 when overcast). a is a local correction value (a negative constant in W/m^2, and b_1 as well as b_2 are locally constant in terms of the cloud correction.

This Hungarian model [8] takes the cloud cover into consideration but not the transparency of the atmosphere (q, T_{m} and l_{atm}). A German model [9] accounts for atmospheric conditions. Equation

$$G = I_n \sin(\alpha) A_d\, e^{-B_d T_{\mathrm{m}} z} (1 - a_d N^{b_d}) \tag{22}$$

shows the global solar irradiance and equation

$$S = I_n \sin(\alpha) A_d\, q^{T_{\mathrm{m}} z} (1 - N^{b_d}) \tag{23}$$

presents the direct solar irradiance formula. The parameters a_d, b_d, A_d and B_d are place-dependent values. A typical parameter set is $a_d = 0.72$, $b_d = 3.2$, $A_d = 0.7$ and $B_d = 0.027$. The diffuse solar irradiance can be calculated by a simple subtraction.

$$D = G - S \tag{24}$$

The calculated global solar irradiance is shown in Fig. 5 over a year with weekly resolutions during clear sky conditions in Veszprém, Hungary. The red curve tracks the maximum values of the astronomical noons on each curve (analemma curve). This curve shows the effect of

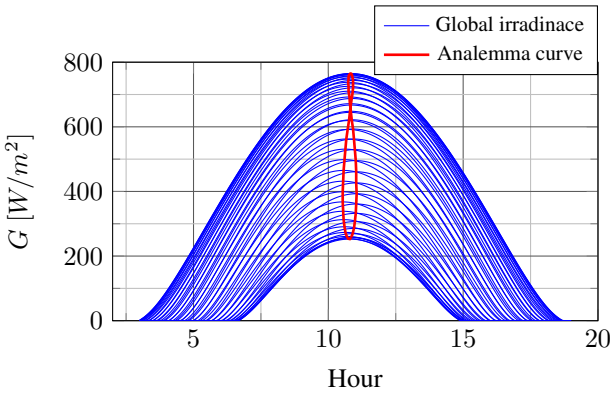


Figure 5: Calculated global solar irradiance on a horizontal surface over the year in Veszprém according to weekly resolutions and the analemma curve (astronomical noon times)

the E_t, the duration of which is approximately 11 hours because the longitudinal position of Veszprém is 17.9° and this causes a delay of about 1 hour (the local time zone is +1).

Fig. 5 clearly illustrates the difference in irradiance between the winter and summer periods.

Irradiance on the slope

For the sake of simplicity the Cartesian coordinate system is used, the axes of which are S (South), W (West) and Z (Zenith). The azimuth angle ψ in the direction S-W-N-E-S and the altitude of the Sun α have been calculated, so the normed vector that points towards the Sun is $\overrightarrow{OS} = [\cos(\alpha)\cos(\psi),\ \cos(\alpha)\sin(\psi),\ \sin(\alpha)]^\top$. If the surface Σ exhibits an angle of inclination of β to the horizon and an orientation of γ, which defines γ as negative, positive or zero if and only if the surface faces SE, SW or S, respectively, then the normed vector of the surface (contained within the same half-space as the Sun) is $\overrightarrow{n}$ and exhibits the altitude $\nu = 90^o - \beta$ so $\overrightarrow{n} = [\cos(\nu)\cos(\gamma),\ \cos(\nu)\sin(\gamma),\ \sin(\nu)]^\top$. The angle between $\overrightarrow{OS}$ and $\overrightarrow{n}$ is

$$\begin{aligned}\theta_{\text{surf}} &= \arccos\Big(\cos(\alpha)\cos(\psi)\cos(\nu)\cos(\gamma)+\\ &\quad+\cos(\alpha)\sin(\psi)\cos(\nu)\sin(\gamma)+\\ &\quad+\sin(\alpha)\sin(\nu)\Big)\\ &= \arccos\Big(\cos(\alpha)\cos(\nu)\Big(\cos(\psi)\cos(\gamma)+\\ &\quad+\sin(\psi)\sin(\gamma)\Big)+\sin(\alpha)\sin(\nu)\Big),\end{aligned} \tag{25}$$

since $\overrightarrow{OS}$ and $\overrightarrow{n}$ are normed. So $\overrightarrow{OS}$ meets Σ at the angle $\alpha_{\text{surf}} = 90^o - \theta_{\text{surf}}$.

Eq. 23 yields the direct solar irradiance on a horizontal surface with solar altitude angle α. The solar altitude

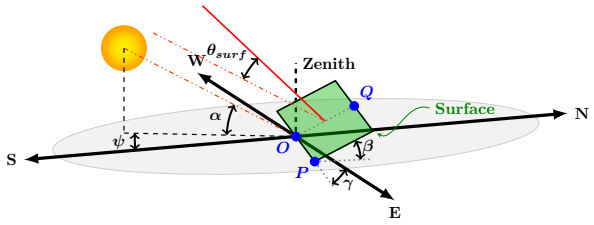


Figure 6: The angle of the solar beam and the zenith angle between the surface and the slope β and orientation γ

angle on the slope is α_{surf}. To calculate the corrected direct solar irradiance, equation

$$S_{\text{surf}} = S\,\frac{\sin(\alpha_{\text{surf}})}{\sin(\alpha)} \tag{26}$$

must be used. Calculation of the diffuse solar irradiance is far from trivial. If the diffuse solar irradiance is consistent with the all-sky, equation

$$D_{\text{surf}} = D\,\frac{1+\cos(\beta)}{2} \tag{27}$$

can be used to estimate the value of the irradiance. The global solar irradiance of the slope is the sum of the direct and diffuse solar irradiances:

$$G_{\text{surf}} = S_{\text{surf}} + D_{\text{surf}}. \tag{28}$$

4. Model verification

The aim of this section is to verify the solar irradiance model proposed in the previous section against basic engineering expectations. Figs. 7 and 8 show the results of the implemented model on 1 March ($d_n = 60$) and 1 July ($d_n = 121$) and the geographical position of Veszprém, Hungary. Both figures present the value of the solar irradiance for various orientations as well as the tilt and azimuth angles.

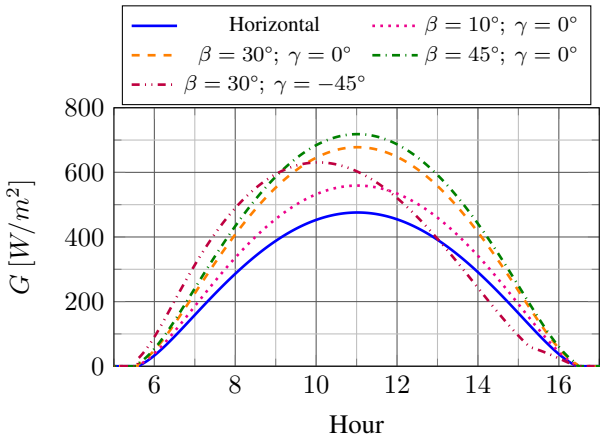


Figure 7: Calculated global solar irradiance on sloping terrain of various orientations on 1 March in Veszprém, Hungary

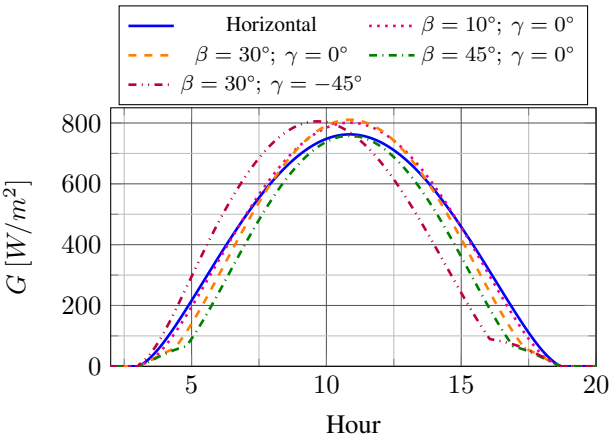


Figure 8: Calculated global solar irradiance on sloping terrain of various orientations on 1 July in Veszprém, Hungary

Irradiance on a horizontal surface is denoted by a blue solid line. The other curves show the value of the solar power with a non-zero tilt angle: 10° (**magenta** dotted line), 30° (**orange** dashed line with an azimuth angle of zero, **purple** dashed dotted line with an azimuth angle of −45° (South-East)) and 45° (**green** dashed dotted line).

The peak values of the curves in Fig. 7 correlate with the tilt angle (β). As the tilt angle increases, so does the irradiance when the azimuth angle is constant ($\gamma = 0°$). As the azimuth angle changes towards the southeast, the peak of the curve shifts towards the sunrise.

The effect of the tilt angle is somewhat different in Fig. 8. On 1 July, the absolute peak can be observed for cases $\beta = 10°$ and $\beta = 30°$ but the irradiance for the whole day is higher if the tilt angle is smaller. This effect is caused by the angle of the solar altitude (smaller zenith angle). If the tilt angle is greater than the minimum horizontal zenith angle, the solar irradiance will decrease on the slope. On 1 March the maximum angle of the solar altitude did not reach 45°. In other words, a greater tilt angle generates more power on sloping terrain. The effect of the azimuth angle can be seen in Fig. 7.

5. Conclusions

A solar irradiance model is presented in this work that can be used for forecasting the solar energy absorbed by an inclined surface, e.g. a PVPP. The developed model is preliminary in nature and serves as a basis for further developments, e.g. forecasting the power of the solar energy under cloudy conditions. The novel element of the model is that the relative path length of the solar beam in the atmosphere can be calculated more accurately. The calculated irradiance an inclined surface is exposed to is another novel element. The proposed model has been verified against engineering expectations and the results show that it works well.

The next step will be the validation of the model using real measured parameters and meteorological data. By extending the model with an electrical power module, it will also be possible to forecast the amount of electrical energy produced.

Acknowledgement

We acknowledge the financial support of Széchenyi 2020 under the EFOP-3.6.1-16-2016-00015. We acknowledge the financial support of Széchenyi 2020 under the GINOP-2.2.1-15-2017-00038. A. Magyar was supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences.

REFERENCES

[1] Braun, J.E.; Mitchell, J.C.: Solar geometry for fixed and tracking surfaces, *Sol. Energy*, 1983 **31**(5), 439–444, DOI: 10.1016/0038-092X(83)90046-4

[2] Bird, R.E.; Riordan, C.: Simple solar spectral model for direct and diffuse irradiance on horizontal and tilted planes at the earth's surface for cloudless atmospheres, *J. Clim. Appl. Meteorol.*, 1986 **25**(1), 87–97, DOI: 10.1175/1520-0450(1986)025%3C0087:SSSMFD%3E2.0.CO;2

[3] Danandeh, M.A.; Mousavi G., S.M.: Solar irradiance estimation models and optimum tilt angle approaches: A comparative study, *Renew. Sust. Energy Rev.*, 2018 **92**, 319–330, DOI: 10.1016/j.rser.2018.05.004

[4] Li, D.H.W.; Chau, T.C.; Wan, K.K.W.: A review of the CIE general sky classification approaches, *Renew. Sust. Energy Rev.*, 2014 **31**, 563–574, DOI: 10.1016/j.rser.2013.12.018

[5] Akarslan, E.; Hocaoglu, F.O.; Edizkan, R.: Novel short term solar irradiance forecasting models, *Renew. Energy*, 2018 **123**, 58–66, DOI: 10.1016/j.renene.2018.02.048

[6] Iqbal, M.: An introduction to solar radiation (Academic Press, London, UK), 1st edn., 1983, DOI: 10.1016/B978-0-12-373750-2.X5001-0, ISBN: 978-0-12-373750-2

[7] Duffie, J.A.; Beckman, W.A.: Solar engineering of thermal processes (John Wiley & Sons, New Jersey, USA), 2013, DOI: 10.1002/9781118671603 ISBN:978-0-47-087366-3

[8] Práger, T.; Ács, F.; Baranka, G.; Feketéné Nárai, K.; Mészáros, R.; Szepesi, D.; Weidinger, T.: A légszennyezõ anyagok transzmissziós szabványainak korszerûsítése III. fázis Részjelentés 2., 1999

[9] Kasten, F.: Strahlungsaustausch zwischen Oberflächen und Atmosphäre, *VDI-Bericht*, 1989 (Nr. 721. S.), 131–158

Appendix: Nomenclature

Symbol	Description	Value/dimension
E_0	Reciprocal of the relative Sun-Earth squared distance	$(r_0/r)^2$
r_0	Mean Sun-Earth distance (Astronomical Unit: [AU])	1 AU
r	Actual Sun-Earth distance in AU	[AU]
d_n	The day of the year	1-365
I_0	Solar constant	1367 W/m^2
I_n	Corrected solar constant	$I_0 E_0$
δ	Sun's declination	$\pm 23.5°$
E_t	Equation of time	$\sim \pm 15$ min
t_{UTC}	The time in UTC	0-24
ϕ_{lat}	Geographic latitude	$\pm 90°$
ϕ_{long}	Geographic longitude	$\pm 180°$
α	The angle of the solar altitude	0°-90°
θ_{z}	Solar zenith angle	$90° - \alpha$
ψ	Solar azimuth angle	$\pm$ 180°
h	The solar hour angle	0°-360°
q	Clear sky transmissivity	0.93
T_{m}	Linke turbidity factor	1.8-4 in Central EU
z	Relative solar beam length in the atmosphere	
l_{atm}	Length of the path of the solar beam through the atmosphere	[km]
R_{Earth}	Median radius of the Earth including the atmosphere	$\sim$ 6470 km
r_{Earth}	Median radius of the Earth from the ground	$\sim$ 6370 km
G	Global solar irradiance	[W/m^2]
N	Cloudy parameter	0-1
A_d	Place-dependent constant	
B_d	Place-dependent constant	
a_d	Place-dependent constant	
b_d	Place-dependent constant	
D	Diffuse solar irradiance	[W/m^2]
S	Direct solar irradiance	[W/m^2]
β	Angle above the surface from the horizontal position	0°–90°
γ	Surface azimuth angle	$\pm 90°$
θ_{surf}	The solar zenith angle on the surface	$90° - \alpha_{\mathrm{surf}}$
α_{surf}	The angle of the solar altitude on the surface	0°-90°
S_{surf}	Direct solar irradiance on the surface	[W/m^2]
D_{surf}	Diffuse solar irradiance on the surface	[W/m^2]
G_{surf}	Global solar irradiance on the surface	[W/m^2]

HUNGARIAN JOURNAL OF
INDUSTRY AND CHEMISTRY
Vol. 47(1) pp. 65–70 (2019)
hjic.mk.uni-pannon.hu
DOI: 10.33927/hjic-2019-10

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

AGGREGATION OF HETEROGENEOUS FLEXIBILITY RESOURCES PROVIDING SERVICES FOR SYSTEM OPERATORS AND THE MARKET PARTICIPANTS

ISTVÁN BALÁZS *[1], ATTILA FODOR[1], AND ATTILA MAGYAR[1]

[1]Department of Electrical Engineering and Information Systems, Faculty of Information Technology, University of Pannonia, Egyetem u. 10, Veszprém, 8200, HUNGARY

Power systems characterized by large, centralized generation sources and the typical flow of energy from the transmission grid to the distribution grid towards consumers are evolving. The increasing penetration of intermittent and distributed renewable energy generation is forcing system operators to increase the volume of balancing capabilities and procure flexibility services at the distribution grid level that must be supported by the aggregation of small-scale resources connected at the distribution grid. This paper suggests an aggregator framework that provides services for both operators of transmission and distribution systems while optimizes its portfolio to perform on wholesale energy trading markets too. Overlaying phases of multi-period optimization runs are proposed that incorporate stochastic renewable energy generation as well as load forecasts and, moreover, the continuously changing business context while enabling cooperation between optimization phases throughout the business process.

Keywords: generation aggregator, optimization, distributed energy resources, TSO-DSO coordination, smart grid

1. Introduction

Conventional power systems are characterized by large generation sources that inject power into the transmission grid, which is transported to distribution networks before being delivered to the end users. Power flows one way from the high-voltage transmission grid to the end user at low-voltage networks. Centralized, dispatchable and predictable generation provides flexibility at the transmission level to the electricity system to balance generation and demand.

The increasing amount of distributed and renewable generation (from around 21% of total net electricity generation in 2010 to 44% in 2030 [1]) has transformed generation into a more variable and intermittent source of energy. Demand has become more active, emphasizing the engagement of consumers. Distributed generation (DG), demand response (DR) and storage facilities will become important components of power systems in the future. These resources are connected to low- and medium-voltage networks, thus, making the distribution grid a crucial element of the electricity sector.

The increasing penetration of intermittent generation and distributed energy resources has already forced TSOs (Transmission System Operator) to increase the volume of balancing capabilities and start procuring services for system balancing not only from the transmission grids but also from distribution grids. An important concept is flexibility, that is the modification of generation injection and/or consumption patterns in reaction to an external signal (price signal or activation), in order to provide a service within the energy system [2]. It is the active management of an asset that can impact the balance of a system or power flows to the grid on a short-term basis. The proper management of available flexibility, both in terms of generation and demand, can help to compensate for the lack of certainty with regard to renewable sources.

On the other hand, DSOs (Distribution System Operator) have started to manage congestion actively in their distribution grids and consider procurement of flexibility services to redispatch the system at the distribution level. As a result, the same flexibility resources could also be potentially used for congestion management and voltage control by the TSO and DSO [3]. Conflicts of interest may arise between the TSO and DSO that must be managed by suitable coordination schemes [4], but this also provides an opportunity for players on the distribution grids to offer flexibility capabilities to multiple customers resulting in higher demand.

An aggregator, that is a new market agent, will play a central role in collecting resources on the distribution grid and involving them in markets that are unavailable for them individually. USEF (Universal Smart Energy

*Correspondence: balazsistvan08@protonmail.com

Framework) Foundation's Aggregator Workstream analyzed the different topics related to the role of the aggregator whilst paying particular attention to demand-response aggregation as well as the relationship between an aggregator and the BRP (Balance Responsible Party)/supplier. Seven different models to implement aggregation were identified, and advantages as well as limitations evaluated [5]. Flexibility resources were investigated in detail and an information model set up by SmartNet D1.2 [6] that contains a mathematical description of the dynamic behaviour of the resource, its constraints in terms of flexibility provision, and a formulation of the different components of costs needed to provide flexibility. The research project BestRES (Best practices and implementation of innovative business models for Renewable Energies aggregators) [7] explored different ways in which an aggregator can create value, and categorized services that are provided into internal (own balancing) as well as external reasons (wholesale, retail, reserve capacity mechanisms).

An aggregation function manages distributed energy resources, the optimization of dispatch by taking into consideration the states of a time-variant system is a core capability. With regard to isolated microgrids, Olivares et al. [8] proposes a stochastic-predictive control approach. The uncertainty of an isolated grid for a single purpose (generation and load balancing) is addressed, a two-stage stochastic receding horizon approach is presented that can be generalized to develop formulations for market-connected multi-purpose optimization scenarios.

The objective of this study is to develop a framework for an aggregator that implements the aggregator role of heterogeneous distributed energy resource management, complies with the typical electricity market model, follows its business processes, and operates in both wholesale energy and electricity balancing markets.

The present paper provides novelties in terms of aggregator roles. It is based on a centralized approach where a new market agent, the aggregator, manages the flexibilities of a portfolio and offers services to TSO, DSO and external BRPs. In addition, it participates in wholesale energy trading activities and co-optimizes resource portfolios for both energy trading as well as flexibility services. The building blocks of such an aggregator have been determined: 3 phases of an overlaying optimization process are recommended, optimization requirements are defined and input/output parameters determined to enable cooperation between optimization phases throughout the business process.

Section 1 introduced challenges that concern the energy sector, defined a problem statement and presented current research as well as implementations. Section 2 describes the architecture of an aggregator extending the virtual power plant concept with heterogeneous resources of flexibility and the optimization module. Differences between typical aggregator definitions and functions proposed in this paper are highlighted. An architecture is planned where optimization is separated from the generator control module as it has to solve a much more complex problem than regular, built-in dispatch functions face. The business context is presented that assumes a complex optimization approach as introduced in Section 3. Here optimization requirements are collected to fulfil the objectives of previous sections and to optimize process phases, data exchange between phases is determined.

2. Aggregator framework

The flexibility center (FC) is introduced in this paper as an entity that implements the role of the aggregator. It is a framework of tools and functions that enable the aggregation of distributed, heterogeneous flexibility resources and provide services in wholesale markets. Components of the FC will be presented as well as its operating environment in order to provide a framework for optimization of heterogeneous energy resources.

2.1 Aggregator role

According to the definition derived from the Universal Smart Energy Framework (USEF) [9], an aggregator is responsible for acquiring flexibility from prosumers, aggregating it into a portfolio, creating services that draw on the accumulated flexibility, and offering these flexibility services to different markets that serve various market players. In return, the aggregator receives the value it creates on these markets and shares it with the prosumer as an incentive to shift its load. In the definition above, USEF only restricts the scope of the aggregator for prosumers by limiting the activities of the aggregator with regard to demand response aggregation. The EU Commission proposal for the recast of the E-Directive [2] defines the aggregator as a market participant that combines multiple customer loads or generated electricity for sale, purchase or auction in any organised energy market. In this study the Commission's definition is used that incorporates aggregation of all types of decentralized energy resources.

The objective of the research is to identify optimization use cases of an aggregator, so assumptions have to be made to simplify the market environment the aggregator operates in. To be able to ignore challenges concerning the dissociation of energy supply / wholesale energy trading and flexibility activation, which is a significant change to current market models, the FC implements the role of the aggregator using an integrated aggregator model [5]. Following the integrated model, the roles of the supplier/trader and aggregator are combined into one market party, both of which are part of the same balancing group, contracted by the same BRP. Since compensation for imbalances and the open supply position are unnecessary, portfolio optimization efforts rather than market issues remain the primary focus.

Definitions and implementations of an aggregator, as referenced in Section 1, outline a model aggregator as a

service provider that responds to external requirements. In this study, the role of the aggregator is extended to operate on the wholesale energy markets as well, in cooperation with its BRP/energy trader. It is proposed that the Flexibility Center should both fulfil external requests and participate in energy trading to optimize economic gain.

2.2 Architecture

The proposed Flexibility Center contains individual modules that cooperate to implement the roles of aggregation.

The SCADA (Supervisory Control and Data Acquisition) system is responsible for data acquisition and control. It can be considered as a low-level interface to the controlled resources. Standard building block.

The AGLC (Automatic Generator Loading Control) module controls power levels concerning elements of the portfolio using SCADA services based on the expected optimized dispatch. Standard building block.

The forecast module provides forecasts with regard to generation, consumption volumes and market prices. Generation and consumption is forecast as a standard function for any implementations that control intermittent resources. However, price forecasting is an additional function required by the trading responsibility of the Flexibility Center.

The proposed architecture of the aggregator defines optimization as an individual module. According to the proposed architecture, it has an extended responsibility, enabling the Flexibility Center to control a diverse portfolio of distributed resources and optimize their capacities in order to both participate in wholesale energy markets and fulfil external requests for flexibility activation.

2.3 Services offered and customers

The Flexibility Center provides services for its customers in terms of aggregating the flexibility capabilities of distributed resources. In terms of the FC, the reason why a customer requested flexibility is irrelevant. However, to prepare high-quality services that meet the technical requirements of the customer and are profitable for the FC, a set of offered services has to be defined. Fig. 1 presents connections between flexibility providers, the Flexibility Center and customers. The Flexibility Center provides services to potential customers:

Internal BRP: FC provides optimal dispatch of the portfolio and trading recommendations to maximize profit on the day-ahead and intraday energy markets. FC can use the portfolio of its own balancing group to minimize imbalance costs of providing such a service for the balance responsible party (BRP) of the balancing group near the delivery time.

TSO: TSO procures balancing reserves to ensure resources for load-frequency control. FC can bid on both balancing capacity and balancing the energy market by offering frequency restoration reserves (FRR).

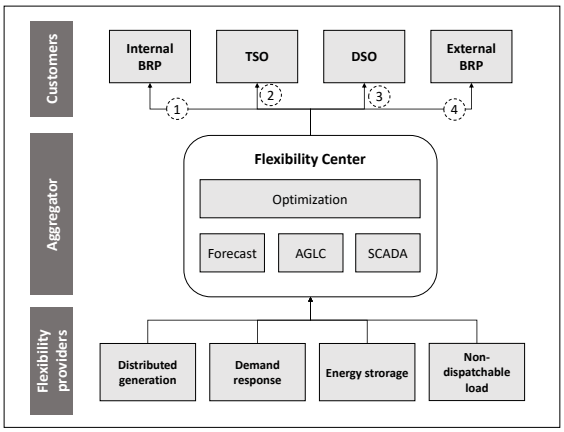


Figure 1: Flexibility Center services and operational context

DSO: DSO's will be empowered to actively purchase local flexibility capabilities to mitigate voltage and grid congestion problems due to the increasing penetration of intermittent and distributed energy resources in the distribution system. FC can bid on future local flexibility markets by offering services for congestion management in the distribution network [9]. Voltage / reactive power controls may also be a positive use case, but these are not considered in this study.

External BRP: Similarly to the internal BRP service, it is also economically worthwhile to balance an external BRP portfolio [10].

2.4 Flexibility resources

Distributed generation refers to power generation facilities connected to the distribution grid. It may consist of predictable, dispatchable and intermittent generation. Demand response is defined as the changes in energy use by end users (domestic and industrial) from their current/normal consumption patterns in response to market signals. Energy storage can be used to store an excess of energy during high generation periods and transform it back into electricity during periods of high demand or to balance the power system. A non-dispatchable load is the consumption of an end user that does not participate in demand response programs.

3. Optimization phases

Markets, in which the FC operates, change over time during the same delivery period. Information on the market, assets of the portfolio, contractual commitments and weather conditions all change over time and the FC optimization function has to reflect such changes. An optimization approach is proposed that takes into consideration the changing objectives and status of the optimization phases.

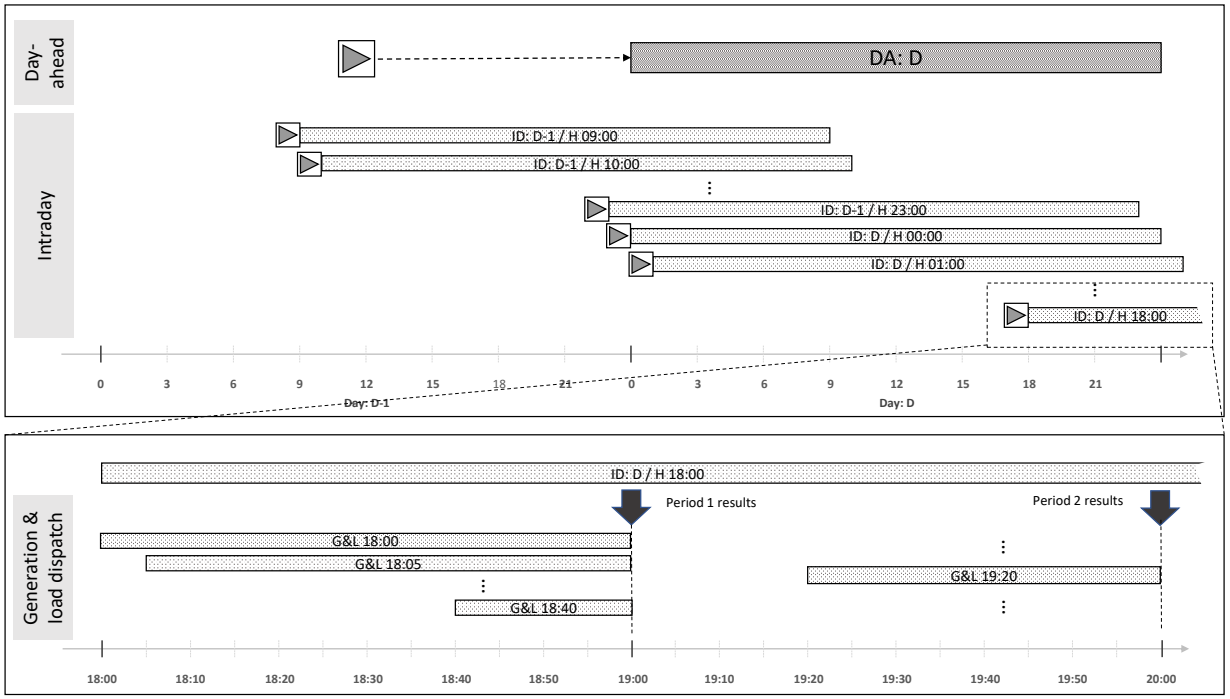


Figure 2: Examples of optimization phases and look-ahead windows

3 phases are identified that are built upon each other. Each phase shares the same primary objective to maximize profit, but the objective function, set of variables, parameters and constraints may differ as the time of delivery approaches and former calculations/forecasts become reality. In order to prepare a consistent optimization framework for the FC, optimization phases are identified and descriptions provided.

Fig. 2 depicts an example of overlaying optimization runs. Day-ahead planning is run once a day, intraday planning is initiated every hour, and Generation and load dispatch is executed every 5 minutes. Once initiated, all optimization runs have to update optimization parameters so they can provide progressively more precise results.

3.1 Requirements of optimization

Requirements of wholesale energy and flexibility activation call for a complex optimization approach. In this paper, common optimization requirements are identified, however, detailed formulations will be developed as part of future research.

The Flexibility Center as an aggregator manages distributed resources that use the grid but does not operate the grid. It is not responsible for nor able to perform network monitoring and control, thus, the grid is not modeled, rather an economic dispatch problem is solved by FC optimization. Due to uncertainties concerning consumption and intermittent renewable generation, a stochastic optimization function has to manage multiple, probability-weighted load and generation scenarios. Energy storage, ramping and unit commitment considerations require a multi-period planning horizon.

Objective functions are formulated to maximize profit and optimization phases share execution results, but they all solve a different optimization problem. Variables, parameters, cost functions and constraints must be aligned accordingly.

Contractual commitments have no cancellation value, these are obligatory constraints. Orders from contracted customers (TSO, DSO, BRP) must be superposed to fulfil expectations and incorporated into the portfolio optimization model.

3.2 Day-ahead planning

The goal of the day-ahead planning phase is to provide portfolio dispatch for schedule nominations containing make-or-buy recommendations based on market price forecasts. Since this is performed on a daily basis, the run time is not critical. Multi-stage stochastic unit commitment and economic dispatch optimization are recommended to mitigate generation and load uncertainties. Periods of 15 minutes in duration are necessary to meet the requirements of the nomination process.

Input parameters for day-ahead planning are reserved flexibility capacities, contractual commitments (e.g. reserved capacities, commercial contracts), the state of the system at the beginning of day D according to the previous intraday run, technical and price parameters with regard to the assets of the portfolio, the generation forecast of intermittent resources, the consumption forecast of loads, and the day-ahead spot price forecast. The out-

Table 1: Timeframes of phases

Phase	Interval	Number of runs	Periods
DA	1 calendar day	1 / day	96×15 min periods
ID	24 hours	24 / day	96×15 min periods
G&L	$60-5$ mins	12 / hour	$12-1 \times 5$ min periods

put contains the power dispatch of the portfolio including binary unit commitment decisions and day-ahead trading recommendations.

Table 1 summarizes optimization intervals and the duration of each phase. The results of day-ahead optimization are inputted into schedule nominations and day-ahead trading so a calendar day is covered but executed according to nomination schedules from the middle of the previous day.

3.3 *Intraday operation*

The primary objective of the intraday run is to determine the system state at the end of the first hour which is then used by generation and load dispatch as a target state (19:00 according to the example in Fig. 2). It can also be used to fine-tune day-ahead results and produce inputs for intraday trading and schedule modifications.

Intraday optimization takes into consideration unit commitment decisions, nominated day-ahead as well as intraday schedules, reserved capacities, more precise weather as well as price forecasts, portfolio parameters and the current state of the system provided by the SCADA module, price parameters of the assets from the portfolio, and activation orders from customers. It is executed hourly by applying a 24-hour look-ahead window. The run-time is critical and the unit commitment is not revised, but probability-weighted generation and load scenarios still need to be generated for stochastic optimization. Multi-period stochastic optimization is recommended. The duration of periods is 15 minutes to support intraday trading.

3.4 *Direct generation and load dispatch*

Intraday optimization performs a 24-hour run and determines a target system state for a direct generation and load dispatch run to calculate optimized dispatch for the Automatic Generator Loading Control module. Here the optimization interval is 1 hour and the duration of calculations is 5 minutes. The result of the first period is used by the AGLC to directly control the assets of the portfolio.

The initial parameter set is the actual system state provided by the SCADA module, the target state is determined by the intraday run. It also follows the shrinking horizon model by being executed every 5 minutes and performing calculations up until the hourly target. Results must be sent to AGLC to control assets since the run-time is extremely critical. A deterministic optimization linear programming formulation is proposed.

4. Conclusion

In this paper, an aggregator framework was presented that aims to manage distributed energy resources, provide services for system operators, support BRPs to minimize imbalances in the balancing group, and determine energy trading volumes in day-ahead and intraday markets. 3 phases of overlaying process optimization were recommended in line with a regular marketing process.

A. Appendix: Acronyms

Abbr.	Description
AGLC	Automatic Generator Loading Control
BRP	Balance Responsible Party
BSP	Balancing Service Provider
DG	Distributed Generation
DR	Demand Response
DSO	Distribution System Operator
FC	Flexibility Center
SCADA	Supervisory Control and Data Acquisition
TSO	Transmission System Operator

B. Appendix: Definitions

Balancing Group: a group of market participants (consumers, producers, traders) who optimize costs by netting deviations (imbalances) and reduce overall deviations between the projected and reported electricity usage.

Balance Responsible Party: a chosen representative of a balancing group who is responsible for the imbalance of the group.

Balancing Service Provider: a market participant providing balancing services to the TSO.

Balancing Services: (1) balancing energy is energy used by TSO to perform balancing and (2) balancing capacity, namely a volume of capacity that a BSP has agreed to hold to and in respect of to which the BSP has agreed to submit bids for a corresponding volume of balancing energy to the TSO.

Balancing: balancing encompasses all actions and processes through which TSO ensures the system frequency is within a predefined stability range and complies with the amount of reserves needed with respect to the required quality.

Distribution System Operator: responsible for providing and operating low-, medium- and high-voltage networks for the regional distribution of electricity as well as for the supply of lower-level distribution systems and directly connected customers.

Transmission System Operator: responsible for providing and operating high- and extra high-voltage networks for the long-distance transmission of electricity as well as for the supply of lower-level regional distribution systems and directly connected customers.

Prosumer / active customer: customers who consume, store or sell electricity generated on their premises, or participate in demand response or energy efficiency schemes provided that these activities do not constitute their primary commercial or professional activity.

Demand response: the change of electricity load by final customers from their normal or current consumption patterns in response to market signals, including time-variable electricity prices or incentive payments, or in response to acceptance of the final customer's bid.

REFERENCES

[1] Merino, J.; Gómez, I.; Turienzo, E.; Madina, C.: Ancillary service provision by RES and DSM connected at distribution level in the future power system, Tech. Rep., SmartNet project D, 2016, http://smartnet-project.eu/wp-content/uploads/2016/12/D1-1_20161220_V1.0.pdf

[2] European Commission: Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on common rules for the internal market in electricity, 2017

[3] de Jong, G.; Franz, O.; Hermans, P.; Lallemand, M.: TSO-DSO data management report, Tech. Rep., CEDEC, EDSO, ENTSO-E, EURELECTRIC, GEODE, 2016

[4] Gerard, H.; Rivero, E.; Six, D.: Basic schemes for TSO-DSO coordination and ancillary services provision, Tech. Rep., SmartNet project D, 2016, http://smartnet-project.eu/wp-content/uploads/2016/12/D1.3_20161202_V1.0.pdf

[5] de Heer, H.; van der Laan, M.: Recommended practices and key considerations for a regulatory framework and market design on explicit Demand Response, Tech. Rep., Universal Smart Energy Framework, 2017

[6] Le Baut, J.; Leclerq, G.; Viganò, G.; Degefa, M.Z.: Characterization of flexibility resources and distribution networks, Tech. Rep., SmartNet project D, 2017, http://smartnet-project.eu/wp-content/uploads/2017/05/D1.2_20170522_V1.1.pdf

[7] Verhaegen, R.; Dierckxsens, C.: Existing business models for renewable energy aggregators, 2016 , 2689723 http://bestres.eu/wp-content/uploads/2016/08/BestRES_Existing-business-models-for-RE-aggregators.pdf

[8] Olivares, D.E.; Lara, J.D.; Cañizares, C.A.; Kazerani, M.: Stochastic-predictive energy management system for isolated microgrids, *IEEE Transactions on Smart Grid*, 2015 **6**(6), 2681–2693, DOI: 10.1109/TSG.2015.2469631

[9] Framework, U.S.E.: USEF: The framework explained, Tech. Rep., Universal Smart Energy Framework, 2015

[10] Olivella-Rosell, P.; Lloret-Gallego, P.; Munné-Collado, I.; Villafafila-Robles, R.; Sumper, A.; Ottessen, S.Ø.; Rajasekharan, J.; Bremdal, B.A.: Local Flexibility Market Design for Aggregators Providing Multiple Flexibility Services at Distribution Network Level, *Energies*, 2018 **11**(4), 822, DOI: 10.3390/en11040822

HUNGARIAN JOURNAL OF
INDUSTRY AND CHEMISTRY
Vol. 47(1) pp. 71–77 (2019)
hjic.mk.uni-pannon.hu
DOI: 10.33927/hjic-2019-11

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

COMPARISON OF THE APPROXIMATION METHODS FOR TIME-DELAY SYSTEMS: APPLICATION TO MULTI-AGENT SYSTEMS

ÁRON FEHÉR *[1,2] AND LŐRINC MÁRTON[1]

[1]Department of Electrical Engineering, Sapientia Hungarian University of Transylvania, C.1., Târgu Mureş, 547367, ROMANIA
[2]Department of Mathematics, University of Pannonia, Egyetem u. 10, Veszprém, 8200, HUNGARY

This paper presents a review of dominant pole and model-approximation algorithms for delayed systems that can be applied to multi-agent systems. A novel algorithm is proposed to determine an approximation method for multi-agent systems in the platoon configuration with a communication delay. Simulations are presented to show the applicability of the proposed algorithm.

Keywords: time delay, differential equations, asymptotic properties, multi-agent systems

1. Introduction

A Multi-Agent System (MAS) consists of multiple active agents (e.g. vehicles), passive agents (e.g. obstacles), in addition to cognitive agents and their environment [1]. Every active agent is at least partially autonomous and uses a distributed control algorithm [2]. The agents can communicate with each other. This communication structure is defined by a graph, where each vertex corresponds to one agent and each edge to a communication direction.

A platoon is a special MAS configuration, with a linear communication graph. The leading (first) agent implements a reference tracking algorithm. Every other agent implements a consensus with the adjacent agents [3].

With the increase in the number of agents and the physical distance between the agents, the communication delay cannot be neglected. While the stability of such systems is ensured by the consensus protocol [4], the delay will influence the transient behavior of the MAS [5].

The goal of this paper is to compare the existing approximation methods for the transient behavior analysis of MAS with communication delays and present a novel analysis method which can be applied to any MAS system which satisfies a smallness delay condition.

2. Modelling of MAS

A MAS is considered with agents that exhibit single-integrator dynamics [6]. The state-space model of an agent becomes $\dot{x}_i(t) = u_i(t)$, where $x_i \in \mathbb{R}$ denotes the state of the ith agent and $u_i \in \mathbb{R}$ represents the input, $i = 1, 2, \ldots, n$.

A MAS has an underlying communication graph, in which the vertex is an agent and the edge a communication path [7], so the agent ith in the system is the vertex v_i. Let N_i be the set of neighbors of v_i, so that N_i contains all vertices that are connected to v_i.

Consensus algorithm The consensus problem of a MAS is the procedure of gathering every state from the initial condition to a common steady-state. If the communication graph is connected, the consensus with regard to an agent can be reached with the input (*consensus protocol*)

$$u_i(t) = \sum_{j \in N_i} \left(x_j(t) - x_i(t) \right). \tag{1}$$

The adjacency matrix $A = (a_{ij}) \in \mathbb{R}^{n \times n}$ of a graph with n nodes is defined as

$$a_{ij} := \begin{cases} 1, & \text{if } i \neq j \text{ and } v_i \text{ are adjacent to } v_j \\ 0, & \text{otherwise} \end{cases}. \tag{2}$$

The degree matrix $D = (d_{ij}) \in \mathbb{R}^{n \times n}$ of a graph with n nodes shows the number of neighbors for each vertex and can be defined as

$$d_{ij} := \begin{cases} \deg(v_i), & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}, \tag{3}$$

where $\deg(v_i)$ denotes the degree or the number of edges incident to vertex i.

*Correspondence: fehera@ms.sapientia.ro

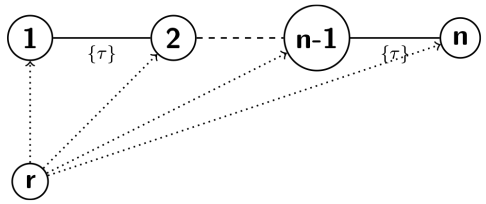


Figure 1: The communication topology of a vehicle platoon system consisting of n vehicles with time delay τ in the communication graph. The dashed line symbolizes more nodes in between, while the dotted line represents the reference input of the nodes.

With this notation the dynamics of the MAS with the consensus protocol is given by

$$\dot{\underline{x}}(t) = -L\underline{x}(t),\ \underline{x}(0) = \underline{x}_0, \tag{4}$$

where L denotes the Laplacian matrix [8] which is constructed as $L = D - A$, D stands for the degree matrix, A represents the adjacency matrix of the graph, $\underline{x} = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix}^\top \in \mathbb{R}^n$ denotes the state vector consisting of the n states of the MAS, and $\underline{x}_0 \in \mathbb{R}^n$ is a constant vector of the initial states.

According to [9] the eigenvalues of a MAS consisting of n agents with a connected communication graph can be ordered as

$$0 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_n. \tag{5}$$

The steady states (equilibria) of the MAS $\underline{x}_{ss}$ are the elements of the null space of L. Given, according to the definition of the Laplacian matrix, $\sum_{j\in N_i} l_{ij} = 0$ [6] for every solution $\underline{x}$ of (Eq. 4): $\lim_{t\to\infty} \underline{x}(t) = \underline{x}_{ss} = \frac{1}{n}\sum_{i=1}^{n} x_i(0)\mathbf{1}$, where $\mathbf{1} = \begin{pmatrix} 1 & 1 & \dots & 1 \end{pmatrix}^\top \in \mathbb{R}^n$.

MAS with delayed communication In the case of delayed communication, the consensus protocol is of the form

$$u_i(t) = \sum_{j\in N_i} \left(x_j(t-\tau) - x_i(t)\right), \tag{6}$$

where $\tau \geq 0$ denotes the constant communication delay which is present among adjacent agents.

According to Refs. [4] and [10], the MAS with communication delay (also referred to as Multi-Agent System with Delays (DMAS)) is

$$\begin{aligned} \dot{\underline{x}}(t) &= -D\underline{x}(t) + A\underline{x}(t-\tau) + \underline{R}, \\ \underline{x}(\theta) &= \underline{x}_0 \in \mathbb{R}^n,\ \theta \in [-\tau, 0], \end{aligned} \tag{7}$$

3. Platoon of vehicles

A platoon of vehicles is a special class of MAS with a communication topology shown in Fig. 1. The kinematic model of a vehicle can be written as

$$\dot{x}_i(t) = u_i(t), \tag{8}$$

where $x_i(t)$ denotes the position of the vehicle and $u_i(t)$ represents the control signal in the form

$$\begin{cases} u_1(t) = k_{p_1}(r - x_1(t)) \\ u_i(t) = \xi_i(x_{i-1}(t-\tau) - d - x_i(t)), \forall i = 2, ..., n \end{cases} \tag{9}$$

where $i \in (1, n)$, $n \in \mathbb{N}, n > 1$, d denotes the prescribed inter-vehicle distance, $\xi_i > 0$, and $k_{p_i} > 0$ are constant control gains. r is the position reference of the first vehicle.

Eqs. 8 and 9 can be written as a system of differential equations with delay:

$$\dot{\underline{x}}(t) = -\Phi\underline{x}(t) + \Gamma\underline{x}(t-\tau) + \underline{f}_1 r - \underline{f}_2 d, \tag{10}$$

where

$$\Phi = \mathrm{diag}\{\begin{pmatrix} k_{p_1} & \xi_2 & \dots & \xi_{n-1} & \xi_n \end{pmatrix}\} \tag{11}$$

denotes the degree matrix with the state feedback,

$$\Gamma = \begin{pmatrix} 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \xi_2 & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \xi_{n-1} & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & \xi_n & 0 \end{pmatrix} \tag{12}$$

represents the adjacency matrix, and

$$\underline{f}_1 = \begin{pmatrix} k_{p_1} & 0 & \dots & 0 \end{pmatrix}^\top, \tag{13}$$

$$\underline{f}_2 = \begin{pmatrix} 0 & \xi_2 & \dots & (n-1)\xi_n \end{pmatrix}^\top. \tag{14}$$

The homogeneous part of the relation in Eq. 10 is of the same form as the relation in Eq. 7, and the term $\underline{f}_1 r - \underline{f}_2 d$ represents the reference as well as inter-vehicle distance induced inflows.

4. Existing methods for the approximation of time-delay systems

In this section, the various current approximation methods for time-delay systems are reviewed. The form of the studied systems is shown by the following relation:

$$\dot{\underline{x}}(t) = -\Phi\underline{x}(t) + \Gamma\underline{x}(t-\tau), \underline{x}(h) = x_0, \tag{15}$$

which is the homogeneous part of Eq. 10, where $\theta(h)$ denotes the initial condition with $h \in [-\tau, 0]$, and has a quasi-polynomial characteristic equation:

$$\lambda I_n + \Phi - \Gamma e^{-\tau\lambda} = 0, \tag{16}$$

where the delay component induces an exponential term.

The roots of Eq. 16 determine the transient behaviour of the system in Eq. 15. As Eq. 15 possesses an infinite number of solutions, it is important to develop such an equivalent system which has a finite number of eigenvalues and is a good approximation of the original system in Eq. 15. If a good delay-free approximation is available, it

could be applied to the transient behaviour analysis and control design for delayed systems.

Let a special solution be denoted by $\tilde{x}$, which is uniquely determined by the value $\tilde{x}(0)$, and independent of $\theta(h), \quad h \in [-\tau, 0)$, thus forming an n parameter family. In a linear autonomous system, this corresponds to the eigensolution generated by exactly n characteristic roots (multiplicities included) which lies in the half-plane $\mathrm{Re}\,\lambda > -1/\tau$, see Ref. [11].

Theorem 1. [11] *Consider a Delay Differential Equation (DDE) system of the form shown in relation Eq. 15 with a Lipschitz criterion of:*

$$(\|\Phi\| + \|\Gamma\|)\tau e < 1, \tag{17}$$

further noted as ***smallness condition****, for every solution* $\underline{x}$ *of Eq. 15 a globally defined solution* $\tilde{x} : \mathbb{R} \to \mathbb{R}^n$ *exists that satisfies the growth condition* $\sup_{t\le 0} \|\tilde{x}(t)\| e^{t/\tau} < \infty$ *such that*

$$\|x(t) - \tilde{x}(t)\| \to 0 \qquad \textit{exponentially as } t \to \infty.$$

For further related discussions, see Refs. [12] and [13]. The aforementioned theorem yields n dominant eigenvalues which can accurately represent a DDE system. The system shown in the relation in Eq. 15 uses the consensus protocol, which ensures that the eigenvalues are located in the left half-plane in the complex region. The final location of the dominant eigenvalues is created as a semicircle with origin 0, radius $1/\tau$, and a negative real part.

4.1 The modified chain approximation

The modified chain approximation method creates an approximating system directly from the state-space representation of the delayed system [14]. For a DDE in the form of Eq. 15, the modified chain approximation method yields an approximating linear system in the form of

$$\begin{aligned}
\dot{\underline{y}}_0(t) &= -\Phi \underline{y}_0(t) + \frac{m}{\tau} I_n \underline{y}_m(t) \\
\dot{\underline{y}}_1(t) &= \Gamma \underline{y}_0(t) - \frac{m}{\tau} I_n \underline{y}_1(t) \\
&\vdots \\
\dot{\underline{y}}_k(t) &= \frac{m}{\tau} I_n \underline{y}_{k-1}(t) - \frac{m}{\tau} I_n \underline{y}_k(t), \quad 2 \le k \le m
\end{aligned} \tag{18}$$

The output vector $\underline{z} = \underline{y}_0$ represents the approximation of the solution $\underline{x}$ of the relation in Eq. 15. The approximating system in the form of a matrix is shown in

$$\dot{\underline{y}}(t) = G\underline{y}(t) \tag{19}$$

with

$$G = \begin{pmatrix}
-\Phi & 0_n & 0_n & \cdots & 0_n & \frac{m}{\tau} I_n \\
\Gamma & -\frac{m}{\tau} I_n & 0_n & \cdots & 0_n & 0_n \\
0_n & \frac{m}{\tau} I_n & -\frac{m}{\tau} I_n & \cdots & 0_n & 0_n \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0_n & 0_n & 0_n & \cdots & \frac{m}{\tau} I_n & -\frac{m}{\tau} I_n
\end{pmatrix} \tag{20}$$

where $\underline{y} \in \mathbb{R}^{mn}$ denotes the state vector, $G \in \mathbb{R}^{(m+1)n \times (m+1)n}$ represents the matrices of the system, $0_n \in \mathbb{R}^{n \times n}$ stands for the zero matrix, n is the number of agents and m denotes the number of approximating equations.

According to Ref. [14] the output of the system (Eq. 19) defined as $\underline{z}(t) = \underline{y}_0(t)$ linearly converges into the solution of the original DDE system such that $c > 0$ and $\sup_{t\ge 0} \|\underline{x}(t) - \underline{z}(t)\| \le \frac{c}{m}$.

4.2 The Lambert W function

The Lambert W function can be used to find the dominant eigenvalues of a quasi-polynomial equation (Eq. 16). Every $W(s)$ function that satisfies

$$W(s)\mathrm{e}^{W(s)} = s \tag{21}$$

by definition is referred to as a Lambert W function [15], where s is either a scalar or matrix complex number function. The Lambert W function has multiple branches denoted as $W_k(s)$ with $k = 0, \pm 1, \pm 2, \ldots, \pm\infty$.

Example 4.1. If a scalar DDE is present in the form of

$$\dot{x}(t) = ax(t) + bx(t-\tau) + cu(t), \tag{22}$$

with $a, b, c, \theta \in \mathbb{R}$, $x(h) = \theta(h)$ for $h \in [-\tau, 0]$, the quasi-polynomial of the homogeneous part can be written as

$$(\lambda - a)\mathrm{e}^{\tau\lambda} = b. \tag{23}$$

If both sides are multiplied by $\tau \mathrm{e}^{-a\tau}$, then

$$(\lambda - a)\tau \mathrm{e}^{\tau(\lambda - a)} = b\tau \mathrm{e}^{-a\tau}, \tag{24}$$

which satisfies Eq. 21 with $W(b\tau \mathrm{e}^{-a\tau}) = (\lambda - a)\tau$, and the eigenvalues can be calculated using the branches of the Lambert W function as

$$\lambda_k = \frac{1}{\tau} W_k(b\tau \mathrm{e}^{-a\tau}) + a. \tag{25}$$

In our case, in terms of the relation in Eq. 15, the aforementioned solution is generalized as

$$\underline{\lambda}_k = \frac{1}{\tau} W_k(\Gamma\tau Q_k) - \Phi, \tag{26}$$

where Q_k can be calculated by solving the equation numerically

$$W_k(\Gamma\tau Q_k)\mathrm{e}^{W_k(\Gamma\tau Q_k)-\Phi\tau} = \Gamma\tau \tag{27}$$

for Q_k [16].

Although many numerical solvers provide native support for the solution of the scalar Lambert W function, the general case requires additional solver tools. Therefore, the LambertWDDE Toolbox [17] was created. The function *find_Sk* assumes τ, Γ and $-\Phi$. The returned values are the eigenvalues $\underline{\lambda}_k$ and the Q_k parameters for a given k branch. The toolbox can create the approximating solution for the given system as

$$\underline{\tilde{x}}(t) = \sum_{k=m_1}^{m_2} \mathrm{e}^{\lambda_k t} C_k^I, \tag{28}$$

where $\tilde{x}$ is the approximating solution of the original delay system and the parameter C_k^I can be computed with the help of *find_CI* for a given $m_1 < k < m_2$ branch. In the scalar case, C_k^I takes the form of

$$C_k^I = \frac{x_0 + b\mathrm{e}^{-\lambda_k\tau}\int_0^\tau \theta(t-\tau)\mathrm{d}t}{1 + b\tau\mathrm{e}^{-\lambda_k\tau}}. \tag{29}$$

4.3 The Quasi-polynomial root-finder algorithm

The quasi-polynomial root-finder algorithm calculates the dominant eigenvalues of a system based directly on the quasi-polynomial equation in Eq. 16.

The quasi-polynomial equation of the system in Eq. 16 can be written as:

$$P(\lambda) = \sum_{k=0}^{N} Q_k(\lambda)\mathrm{e}^{-\alpha_k\tau\lambda}, \tag{30}$$

where Q_k is a polynomial with real coefficients and $\alpha_k \in \mathbb{R}$. The objective is to compute the spectrum in the region of the complex plane $\mathbb{D} \subset \mathbb{C}$ with boundaries $\beta_{\min} < \mathrm{Re}(\mathbb{D}) < \beta_{\max}$ and $\omega_{\min} < \mathrm{Im}(\mathbb{D}) < \omega_{\max}$.

Let the surfaces defined by the real and imaginary parts of $P(\lambda)$ be:

$$\mathrm{Re}(P(\beta,\omega)) = 0 \tag{31}$$
$$\mathrm{Im}(P(\beta,\omega)) = 0 \tag{32}$$

The eigenvalues can be located at the points of intersection of the zero-level curves of the surfaces $\mathrm{Re}(P(\beta,\omega)) = 0$ and $\mathrm{Im}(P(\beta,\omega)) = 0$ as shown in Ref. [18]. The accuracy of the algorithm is increased by Newton's method and by adapting the grid density of $\mathbb{D}$ as shown in Ref. [16].

The Quasi-polynomial root-finder algorithm (QPmR) [19] is implemented in MATLAB. The function expects the region of interest $[\beta_{\min}, \beta_{\max}, \omega_{\min}, \omega_{\max}]$ to be in the complex plane of the polynomial coefficient matrix of the quasi-polynomial where one row corresponds to one polynomial multiplied by the same exponential term. The delay vector, computational accuracy and grid step are also required.

In our case this translates into a region of interest $[-\frac{1}{\tau}, 0, -\frac{1}{\tau}, \frac{1}{\tau}]$ if the smallness condition (Eq. 17) is satisfied. The first row of the matrix of polynomial coefficients contains the coefficients of the delay-free part so that the delay vector is of the form $[0, \tau, 2\tau, ...]$.

The algorithm covers the given region with a mesh grid, then evaluates the quasi-polynomial at each point of the grid by splitting it into a real and an imaginary part. The zero-level curves are then mapped with the help of the *contour* plotting algorithm. The computational error is checked and if it is too large, the algorithm is restarted using a modified grid density as described in Ref. [16]. If the computational error is smaller than the given level of tolerance, the computed dominant eigenvalues are returned.

5. Explicit matrix approximation method

An approximation method was devised where the convergence rate is exponential, the degree of the resulting system in the form of Eq. 4 matches exactly the degree of the delayed MAS given in Eq. 15, and the same properties are exhibited in specific cases.

The Banach fixed-point theorem was used as discussed in Ref. [20] to explicitly find a linear system of the form of Eq. 4 which approximates the homogeneous part of the system in Eq. 10.

If an (X, f) metric space is present and $T : B \to B$ is a contraction with a bounded set $B \subset X$, and $q < 1$ such that

$$f(T(x), T(y)) \leq qf(x, y) \tag{33}$$

by definition T admits a unique fixed point $\tilde{x}$ such as $T(\tilde{x}) = \tilde{x}$, and this fixed point can be found by starting from an arbitrary element $x_0 \in B$ with the sequence

$$x_n = T(x_{n-1}), \tag{34}$$

where $x_n \to \tilde{x}$.

A DDE system is defined in Eq. 15 by the corresponding smallness condition of Eq. 16. Since all normed spaces are metric spaces, the metric space $X = \mathbb{R}^{n\times n}$ is set with f as the induced matrix norm. The contraction $T : B \to B$ is present such that

$$T(\Lambda) = -\Phi + \Gamma\mathrm{e}^{-\Lambda\tau}, \tag{35}$$

and $B = \{\Lambda \in \mathbb{R}^{n\times n} | \|\Lambda\| \leq (\|\Phi\| + \|\Gamma\|)\mathrm{e}\}$. The relation in Eq. 33 holds true for $(\|\Phi\| + \|\Gamma\|)\tau\mathrm{e} < 1$.

Let $\Lambda_1, \Lambda_2 \in B$ such that $\|\Lambda_1\| > \|\Lambda_2\|$. The left side of the inequality in Eq. 33 can be written as

$$\|T(\Lambda_1) - T(\Lambda_2)\| = \|\Gamma\|\|\mathrm{e}^{-\Lambda_1\tau} - \mathrm{e}^{-\Lambda_1\tau}\|.$$

It is evident that $\|\Gamma\| \leq \|\Gamma\| + \|\Phi\|$ and the maximum norm can be used as

$$\|\mathrm{e}^{-\Lambda_1\tau} - \mathrm{e}^{-\Lambda_1\tau}\| \leq \tau\|\Lambda_1 - \Lambda_2\|\mathrm{e}^{\tau\max\{\|\Lambda_1\|,\|\Lambda_2|\}}.$$

It can be seen that

$$\|T(\Lambda_1)-T(\Lambda_2)\| \leq \tau\|\Lambda_1-\Lambda_2\|(\|\Gamma\|+\|\Phi\|)\mathrm{e}^{\tau(\|\Gamma\|+\|\Phi\|)\mathrm{e}}$$

and by applying the smallness condition $\|\Gamma\| + \|\Phi\| \leq 1/(\tau\mathrm{e})$ (Eq. 17),

$$\|T(\Lambda_1) - T(\Lambda_2)\| \leq \|\Lambda_1 - \Lambda_2\|$$

is obtained, which proves that $T : B \to B$ is a contraction. This shows that by solving

$$\Lambda = -\Phi + \Gamma\mathrm{e}^{-\Lambda\tau} \tag{36}$$

for the Λ matrix, a system is created

$$\frac{\mathrm{d}\underline{\tilde{x}}}{\mathrm{d}t} = \Lambda\underline{\tilde{x}}, \tag{37}$$

which approximates the DDE system of Eq. 15. As a result of the proposed iterative method, the eigenvalues and eigenvectors of Eq. 37 approximate, with a given degree of precision, the dominant eigenvalues of Eq. 14.

5.1 Comparison with the existing methods

- Chain approximation:
  - + The result is an approximating system with known system matrices (both eigenvalues and eigenvectors are known).
  - - The resulting system is of a higher degree than the delay system.
  - - The convergence rate of the algorithm is linear.
- Lambert W function:
  - + The result is a trajectory approximation.
  - + The convergence rate of the algorithm is exponential.
  - - The algorithm requires numerical solvers for an exponential matrix equation.
  - - Multiple branches of the Lambert W function must be used to create an accurate approximation, and the number of eigenvalues in a branch cannot be predetermined generally.
- QPmR algorithm:
  - + The algorithm determines the exact number of eigenvalues in the given complex domain, with the given computational error.
  - + The convergence rate of the algorithm is exponential.
  - - The algorithm uses the quasi-polynomial equation, thus, will not contain any information on the eigenvectors.
  - - The algorithm uses numerical solvers to compute the zeros of the zero-level curves created from the quasi-polynomial equation.
- Approximation of an explicit matrix:
  - + The result is an approximating system.
  - + The resulting system is of the same degree as the approximating system.
  - - The algorithm calculates a matrix exponential numerically, which is a compute-intensive task.

6. Simulations and results

Let us consider two cases: a MAS consisting of five and twenty-five agents, respectively, with first-order dynamics in a platoon configuration as shown in the relation of Eq. 10, with the constant initial condition $\underline{\theta}(h) = \underline{x}_0$.

For the comparisons, a 6^{th} order chain approximation was used. In the case of the Lambert W function, the initial matrix $Q_0 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$ and the branches $k = -2, -1, 0, 1$ were used. For the quasi-polynomial root-finder algorithm (QPmR), a symbolic calculation to find the characteristic quasi-polynomial equation of the system was used, and the plane of the search was set to $[-\tau, 0] \times [-\tau j, \tau j]$. The algorithm for explicit matrix approximation was used with the initial matrix $\Lambda_0 = 0_{n\times n}$. The error threshold $1\mathrm{e}{-7}$ was used in every iterative algorithm.

The dominant eigenvalues of the system consisting of five agents, with minor differences, was identified by every approximation method. In the case of the larger system, the chain and explicit matrix approximations could generate a result, while the quasi-polynomial root-finder algorithm and the Lambert W function were determined by numerical calculations. As such, the smaller system was chosen as a point of comparison for the algorithms.

Tables 1 and 2 contain a comparative summary of the four aforementioned algorithms: the number of iterations, the overall computation time and the dominant eigenvalues identified for a platoon consisting of five agents as shown in Fig. 2.

Fig. 3 shows that the linear system is generated by the explicit matrix approximation in just twelve steps for the DMAS that consists of twenty-five agents.

Fig. 4 shows that the resultant approximating system exhibits the same steady-state and transient behavior as the original delayed system using the same initial conditions. Since the initial position falls within the range of

Table 1: The number of iterations and the overall computation time for the algorithms.

Name of algorithm	No. of cycles	Computation time
Chain approximation	1	0.02 s
Lambert W function	35	7.5 min
QPmR algorithm	12	17.18 s
Explicit algorithm	4	0.03 s

Table 2: The eigenvalues returned by the studied algorithms.

Name of algorithm	Eigenvalues
Chain approximation	-0.1080, -0.9471, -1, -2.4736, -4.2511
Lambert W function	-0.1081, -0.9482, -1, -2.4658, -4.1603
QPmR algorithm	-0.1081, -0.9481, -1, -2.4661, -4.1603
Explicit algorithm	-0.1080, -0.9481, -1, -2.4661, -4.1603

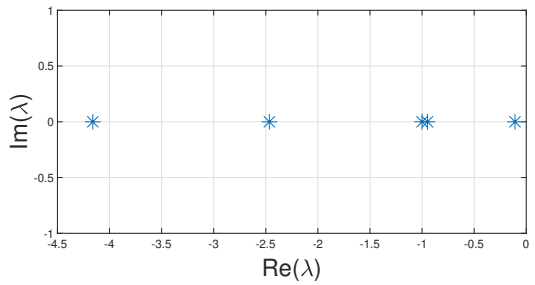


Figure 2: Dominant poles of the DMAS consisting of five agents.

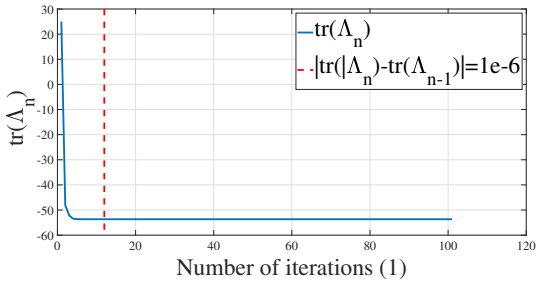


Figure 3: The explicit matrix approximation returns a valid system after 12 iterations for 25 agents.

0 to 100 m, the error of the approximating system is 8 cm in the transient domain and 5 mm under steady-state conditions, as shown in Fig. 5.

7. Conclusions

An iterative algorithm was proposed and tested based on which a degree-preserving approximation model can be created for a class of MAS in a platoon formation consisting of agents that exhibit first-order dynamics in the presence of a small communication delay. The algorithm uses methods of numerical computation. The obtained MAS is of the same degree and steady state as the delayed MAS, moreover, it correctly approximates the transient behavior. The algorithm was compared with existing methods for approximating eigenvalues and systems. Simulations show that the presented algorithm is suitable for the analysis of complex MAS.

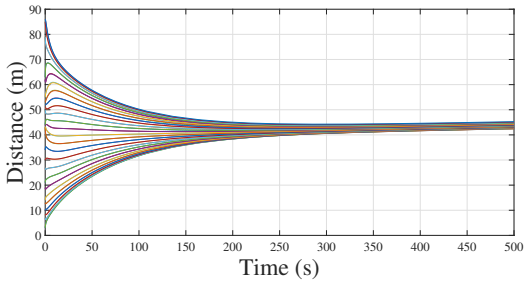


Figure 4: The trajectories of the platoon of the delayed MAS compared to the platoon of the approximated MAS created by the explicit matrix approximation.

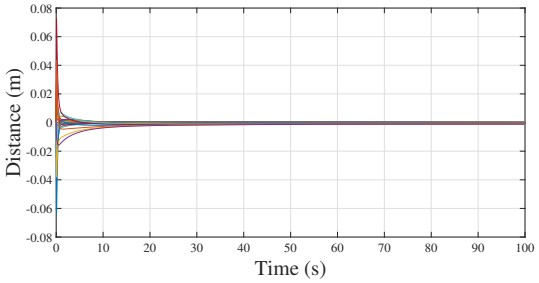


Figure 5: The trajectory of the approximation error.

Acknowledgement

The authors would like to express their gratitude to Dr. Mihály Pituk (University of Pannonia, Hungary) for his useful comments.

REFERENCES

[1] Kubera, Y.; Mathieu, P.; Picault, S.: Everything can be Agent!, in Proceedings of the ninth International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS'2010) (W. van der Hoek, G.A. Kaminka, Y. Lespérance, M. Luck, S. Sen, eds.) (International Foundation for Autonomous Agents and Multiagent Systems, Toronto, Ontario, Canada), 1547–1548

[2] Panait, L.; Luke, S.: Cooperative multi-agent learning: The state of the art, *Auton. Agents Multi-Agent Syst.*, 2005 **11**(3), 387–434, DOI: 10.1007/s10458-005-2631-2, ISBN: 978-981-10-2491-7

[3] Zabat, M.; Stabile, N.; Farascaroli, S.; Browand, F.: The aerodynamic performance of platoons: A final report, Institute of Transportation Studies, Research Reports, Working Papers, Proceedings, Institute of Transportation Studies, UC Berkeley, 1995

[4] Cheng-Lin, L.; Fei, L.: Consensus Problem of Delayed Linear Multi-agent Systems, SpringerBriefs in Electrical and Computer Engineering (Springer Singapore), 2017, DOI: 10.1007/978-981-10-2492-4, ISBN:978-981-10-2491-7

[5] Wim, M.; Silviu-Iulian, N.: Stability and stabilization of time-delay systems: An eigenvalue-based approach (SIAM), 2007, DOI: 10.1137/1.9780898718645

[6] Lewis, F.L.; Zhang, H.; Hengster-Movric, K.; Das, A.: Cooperative control of multi-agent systems: Optimal and adaptive design approaches (Springer), 2014, DOI: 10.1007/978-1-4471-5574-4

[7] Trudeau, R.J.: Introduction to Graph Theory, Dover Books on Mathematics (Dover Publications), 2nd edn., 1994, ISBN: 978-0486678702

[8] Chaiken, S.; Kleitman, D.J.: Matrix tree theorems, *J. Comb. Theory A*, 1978 **24**(3), 377 – 381, DOI: 10.1016/0097-3165(78)90067-5

[9] Mesbahi, M.; Egerstedt, M.: Graph theoretic methods in multiagent networks, Princeton Series in Applied Mathematics (Princeton University Press), 2010, DOI: 10.1515/9781400835355

[10] Cepeda-Gomez, R.; Olgac, N.: An exact method for the stability analysis of linear consensus protocols with time delay, *IEEE T. Autom. Control*, 2011 **56**(7), 1734–1740, DOI: 10.1109/TAC.2011.2152510

[11] Arino, O.; Pituk, M.: More on linear differential systems with small delays, *J. Differential Equations*, 2001 **170**(2), 381 – 407, DOI: 10.1006/jdeq.2000.3824

[12] Győri, I.; Pituk, M.: Asymptotically ordinary delay differential equations, *Functional Differential Equations*, 2005 **12**, 187–208

[13] Győri, I.; Pituk, M.: Asymptotic formulas for a scalar linear delay differential equation, *Electron. J. Qual. Theory Differ. Equ.*, 2016 **1**(72), 1–14, DOI: 10.14232/ejqtde.2016.1.72

[14] Krasznai, B.; Győri, I.; Pituk, M.: The modified chain method for a class of delay differential equations arising in neural networks, *Math. Comput. Model.*, 2010 **51**(5), 452 – 460, DOI: 10.1016/j.mcm.2009.12.001

[15] Corless, R.M.; Gonnet, G.H.; Hare, D.E.G.; Jeffrey, D.J.; Knuth, D.E.: On the Lambert *W* function, *Adv. Comput. Math.*, 1996 **5**(1), 329–359, DOI: 10.1007/BF02124750

[16] Vyhlídal, T.; Lafay, J.F.; Sipahi, R.: Delay systems: From theory to numerics and applications, Advances in Delays and Dynamics (Springer International Publishing), 2013, DOI: 10.1007/978-3-319-01695-5

[17] Yi, S.; Duan, S.; Nelson, P.W.; Ulsoy, A.G.: The Lambert W Function Approach to Time Delay Systems and the LambertW_DDE Toolbox, *IFAC Proceedings Volumes*, 2012 **45**(14), 114–119, DOI: 10.3182/20120622-3-US-4021.00008

[18] Vyhlídal, T.; Zítek, P.: Quasipolynomial mapping based rootfinder for analysis of Time delay systems, in Proc. of IFAC Workshop on Time-Delay Systems, TDS 2003, DOI: 10.1016/S1474-6670(17)33330-X

[19] Vyhlídal, T.; Zítek, P.: QPmR – Quasipolynomial rootfinder: Algorithm update and examples, *Advances in Delays and Dynamics*, vol. 1 (Springer International Publishing, Cham), 2013 299–312, DOI: 10.1007/978-3-319-01695-5_22, http://www.cak.fs.cvut.cz/algorithms/qpmr

[20] Latif, A.: Banach contraction principle and its generalizations, Topics in Fixed Point Theory (Springer International Publishing, Cham), 2014 33–64, DOI: 10.1007/978-3-319-01586-6_2

HUNGARIAN JOURNAL OF
INDUSTRY AND CHEMISTRY
Vol. 47(1) pp. 79–83 (2019)
hjic.mk.uni-pannon.hu
DOI: 10.33927/hjic-2019-12

HUNGARIAN JOURNAL OF INDUSTRY AND CHEMISTRY
HJIC

RULE-BASE FORMULATION FOR CLIPS-BASED WORK ERGONOMIC ASSESSMENT

BENEDEK SZAKONYI *[1], TAMÁS LŐRINCZ[1], ÁGNES LIPOVITS[2], AND ISTVÁN VASSÁNYI[1]

[1]Department of Electrical Engineering and Information Systems, University of Pannonia, Egyetem u. 10, Veszprém, 8200, HUNGARY
[1]Department of Mathematics, University of Pannonia, Egyetem u. 10, Veszprém, 8200, HUNGARY

Modern societies are dominated by computer-based work. As a result, people tend to be seated for most of their working life. Prolonged sedentariness is known to significantly increase the risk of developing unwanted conditions. This paper presents the development of a rule-based expert system module using CLIPS that provides ergonomic assessment. The system was validated by evaluating pre-recorded user logs from real-life office environments. The tests showed that the system is able to perform the required basic assessment functionality, thus, the implementation of more complex features to advance its development is viable.

Keywords: expert system, CLIPS, work ergonomics

1. Introduction

Civilization-related diseases have become more widespread over time, although small improvements to our lifestyle could effectively reduce both their severity and number of sufferers. As adults spend a significant proportion of time working, which for the vast majority involves sitting in front of a computer [1], under what conditions this time is spent is very much relevant.

Thanks to tools of Information Technology and ergonomics experts, it is possible to provide automated solutions, in terms of monitoring and assessing user behaviour, to help more employees avoid the adoption of undesirable and harmful postures whilst at work. Research has already shown that providing such real-time feedback to users can promote this, thus, reducing the impact of related negative effects [2, 3]. But in these cases, wearable sensors were used to track user motion, which can be considered problematic to be applied in real life applications. A possible solution to this might be Microsoft's Kinect sensor [4] that, being a good value-for-money motion capture device, has already been used in similar cases [5, 6]. In spite of this, just as in numerous other papers that concern work-related assessment [7–10], considerably more focus is placed on manual labour where "blue-collar workers" are subject to physically demanding conditions (e.g. production lines, agricultural or construction work, etc.). Even though these fields are of equal importance, such conditions cannot be applied to office environments, where "white-" and "pink-collar workers" spend most of their time, as here the majority of problems does not originate from incorrect movement (e.g. bad techniques when lifting heavy items) but rather from the (utter) lack of movement itself and prolonged sedentariness.

The aim of this research was to develop an expert system module for ergonomic assessment (based on a previously created Lifestyle Coach Framework [11] as shown in Fig. 1 with an emphasis placed on the formulated work ergonomics-based rule set, and to investigate if such a system is capable of properly evaluating user behaviour in office environments.

2. Experimental

2.1 The Framework used

The Lifestyle Coach Framework used in the development process is a rule-based expert system framework using CLIPS (C Language Integrated Production System) [12]. The main reason for using a rule-based solution (instead of neural networks, fuzzy logic, etc.) is the fact that most healthcare experts can express their knowledge using simple IF-THEN-like statements, which is exactly how rule-based systems work.

The CLIPS runtime is responsible for providing the basic functionality of evaluating the current state of the system and deducing the most suitable response and reactions. The statements that describe the current state and different events are called *facts*, while *rules* convey the relevant information concerning what actions to take in

*Correspondence: benedek.szakonyi@virt.uni-pannon.hu

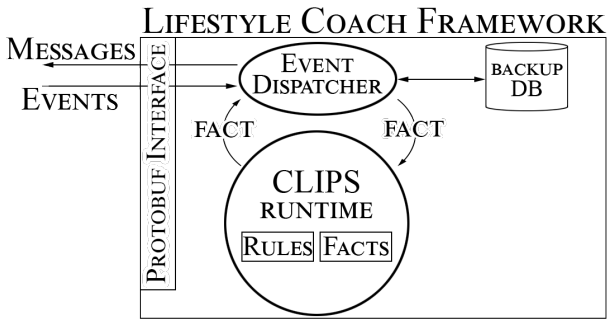


Figure 1: The components of the framework used.

the event (or absence) of such facts, in the form of "simple" IF-THEN expressions. The majority of these components are neither predefined nor hardcoded in the framework itself (except for those that are automatically generated) but are to be defined in light of the specific topic and task at hand in the form of a rule set. Hence the framework can be applied in different fields as long as the problem is interpretable as the evaluation of a log of successive events. These events are transmitted to the system via the Protobuf Interface (that uses Google Protobuf [13]), which also serves as a tool to receive the corresponding responses. The messages of which can be categorized into two types, namely *control* and *event*. The former is used to manage the expert system itself, e.g. to signal that a unit of time has passed and re-evaluation of the state of the system should be commenced. The latter serves as a way of inserting new facts (incoming data) and handling *consequences* (outgoing reactions). How such messages are interpreted and forwarded is the responsibility of the Event Dispatcher. Apart from bridging the gap between the interface and CLIPS runtime, it is also connected to a database that is used to initialize (i.e. it stores the rule-base used), record system behaviour and act as a backup should anything fail.

2.2 Work Ergonomics

To create the rule base required for ergonomic evaluation, methods used to estimate the risks of developing Musculoskeletal Disorders (MSDs) were analysed, with the help of ergonomics experts. The most popular and widely accepted tools used for such tasks are RULA (Rapid Upper Limb Assessment) [14], REBA (Rapid Entire Body Assessment) [15], OWAS (Ovako Working Posture Analysis System) [16] and HARM (Hand Arm Risk-assessment Method) [17].

RULA applies a scoring method for measuring the physical load that workers are subjected to when adopting specific postures by taking into consideration six main body regions: neck, trunk, legs, upper arms, forearms and wrists. REBA works in a similar fashion by using scores but takes into consideration the whole body when such postures are adopted. OWAS uses a 4-digit score for each posture where the digits describe the back, arms, legs and load. HARM accounts for the head, neck, arms and wrists as well as the forces that are applied and the duration over which each posture is adopted. However, these methods cannot be used directly for the matter at hand as originally none of them were intended to be used in regular office environments: RULA was developed to assess work in the textile industry, REBA in healthcare, OWAS in the steel industry, and HARM in hand-intensive professions (e.g. barbers, product assembly/disassembly, woodwork, etc.).

Based on these tools, a basic method of postural assessment was derived as presented in Table 1. It is important to note that when subject to these constraints, the values used were chosen as initial limits as a proof of concept in light of the fact that they may need to be updated. For each body part, the relevant axes of movement were selected and for each axis, ranges were defined. A range can be characterised into three types: appropriate, incorrect and harmful. Appropriate ranges define the desired position the user should take, while harmful ones represent postures that should be avoided or even prohibited. Incorrect ones fall somewhere between the other two categories, when, as far as is feasible, a posture should be prevented since it is considered unhealthy, however, when adopted for short intervals of time it is still acceptable. For the ranges of the inappropriate postures, a frequency is given, that describes that in one work hour, how much time spent in them (in total) is still considered as tolerable.

Apart from evaluating specific body parts based on their position, the time the user spends sitting was also selected to be taken into account, as a general aspect.

2.3 Rule-base Formulation

The formulation of the constraints introduced in Table 1 that are applied to a CLIPS-based rule set could have been determined using numerous different approaches. The one that was selected, builds on the fact that the implementation of these aspects has the same general structure:

1. there are numerical measurement values to be evaluated over each time interval
2. there is a constraint that the measured values are compared to when evaluated
3. there is a frequency that, if exceeded, should trigger a warning to be sent.

Therefore, instead of creating multiple rules for each aspect, one main rule can be defined and applied during the evaluation which can make use of an "aspect template" to check if the related conditions have been adhered to or not. To achieve this, as rules in CLIPS are governed by facts, the aspects are to be in the form of facts. In Fig. 2 the template used for defining aspects is shown. The *aspect_type* serves as a name/identifier for the aspect, while *measurement_type* defines which measurements

Table 1: The advised constraints to be used in the formulated rule base. For each body part, separate ranges were used for the different planes of movement. The acceptable duration within each range was defined (as % of 1 working hour).

Body part	Axis	Range	Frequency
Head/neck	Sagittal	< 0	$< 20\%$
		$0 < \quad < 10$	
		$10 < \quad < 30$	$< 40\%$
		$30 <$	$< 20\%$
	Horizontal	$< \pm 10$	
		$\pm 10 < \quad < \pm 30$	$< 40\%$
		$\pm 30 <$	$< 20\%$
	Frontal	$< \pm 5$	
		$\pm 5 < \quad < \pm 10$	$< 40\%$
		$\pm 10 <$	$< 20\%$
Trunk	Sagittal	< 0	$< 40\%$
		$0 < \quad < 15$	
		$15 < \quad < 25$	$< 40\%$
		$25 <$	$< 20\%$
	Horizontal	$< \pm 5$	
		$\pm 5 < \quad < \pm 10$	$< 40\%$
		$\pm 10 <$	$< 20\%$
	Frontal	$< \pm 5$	
		$\pm 5 < \quad < \pm 10$	$< 40\%$
		$\pm 10 <$	$< 20\%$
Upper arm	Sagittal	< -10	$< 20\%$
		$-10 < \quad < 0$	$< 40\%$
		$0 < \quad < 10$	
		$10 < \quad < 45$	$< 40\%$
		$45 <$	$< 20\%$
	Frontal	< 10	
		$10 < \quad < 20$	$< 40\%$
		$20 <$	$< 20\%$
Forearm	Sagittal	< 60	$< 20\%$
		$60 < \quad < 80$	$< 40\%$
		$80 < \quad < 100$	
		$100 < \quad < 110$	$< 40\%$
		$110 <$	$< 20\%$

should be inspected. In spite of some axes, where ranges of constraints are "evenly distributed" or "symmetrical" (e.g. head frontal, trunk sagittal), the relation between the measured data and the constraint must be provided in the *constraint_type* template, as for other cases, the ranges are "uneven". This type can be either smaller than or equal to, or greater than or equal to, followed by the value to use in the *value_constraint*. The *horizon_minute* slot is used to determine the duration of measurement that is to be examined (e.g. the last 60 minutes). As cases may occur when, despite being warned, users continue to follow their unhealthy lifestyle, a feature that permits repeated warnings of increasing severity (or even more extensive interventions such as turning off the computer screen) could be useful. In the logic developed, checking whether the user has taken the advice given or not is considered to takes less time than the original evaluation interval (this will have an impact on the main evaluating rule, that is detailed later on). With regard to this feature, apart from the *first_occurrence_constraint* that corresponds to the frequency in Table 1 and the *first_reaction_event_code* that defines what actions to take if the limit has been reached,

```
(deftemplateaspect_to_monitor
  (slot aspect_type (type SYMBOL)(default ?NONE))
  (slot measurement_type (type SYMBOL)(default ?NONE))
  (slot constraint_type (type SYMBOL)(default ?NONE))
  (slot value_constraint (type NUMBER)(default ?NONE))
  (slot horizon_minute (type NUMBER)(default ?NONE))
  (slot first_occurrence_constraint (type NUMBER)(default ?NONE))
  (slot first_reaction_event_code (type NUMBER)(default ?NONE))
  (slot repeated_occurrence_constraint (type NUMBER))
  (slot repeated_reaction_event_code (type NUMBER))
  (slot repeated_feedback_horizon (type NUMBER))
)
```

Figure 2: The CLIPS fact template used for describing the structure of ergonomic aspects.

optional *repeated_* slots can be used for this purpose in the template.

As the evaluating CLIPS rule is lengthy, only a short explanation of the logic behind it is provided here. Over each evaluation interval, the number of measurements that exceed the given constraints during the corresponding time horizon is counted, and if this value is above the acceptable limit, a request to send feedback is made and the related measurements are labelled as evaluated (according to the current aspect). If only a few "bad" measurements are identified, the evaluation is considered to be finished, unless repeated feedback is received with regard to the current aspect. In that case, it must be determined if a "first feedback" was given recently, and if it was, the number of measurements that exceed the limit must be compared to the occurrence constraint of the second feedback. If this is exceeded, a second feedback is sent.

2.4 Experiments for Rule-base Validation

As the aim of the research was to provide feedback for users concerning their postures by using a rule base consisting of ergonomic rules, a method for recording and identifying their postures was needed. To accomplish this, a Microsoft Kinect v2 sensor was used. As it is an easy-to-use and relatively small device, it was possible to insert it into the real working setups of the willing participants, with only slight modifications in their environments. The general setup that was created for all users was the following: each user was seated at a desk with a personal computer that consisted of 1 or 3 displays, a keyboard, a mouse and a landline telephone. The sensor was placed on a tripod behind and slightly above the "main" screen (i.e. the one in front of the user). The Kinect was tilted forward in order to ensure as much of the participant as possible was in view (from the top of the head down to the waist/hips).

Users were monitored according to two different "behaviour modes", one being general "everyday" attitude where they completed their daily computer-based tasks as usual. As this inherently meant that users might have had to leave their desks, long measurement sessions were recorded (up to 8 hours in duration) where logging was

suspended once the absence of the participant was detected until their return, when it was resumed (such "gaps", of course, were then taken into account as part of the evaluation). In the other "mode", where considerably shorter measurement intervals were used in order to ensure sessions that definitely consisted of unsuitable behaviour, the users were asked to adopt some evidently inappropriate postures.

As the framework used offers a customisable time unit for defining when an interval stops (i.e. it can last for an hour, a minute, a second, etc.), it was possible to load and evaluate the logs gathered using this method quickly.

3. Results and Discussion

Based on the evaluation method suggested by the ergonomist experts, a total of 11 aspects have been formulated (by using 34 facts), 9 for the positions of the body parts (one for each axis using the constraints shown in Table 1) and 2 regarding sedentariness, where the acceptable duration of continuous sitting was chosen to not exceed 60 minutes (the first aspect was related to sending a warning once this limit had been reached, the second was used to alert when a participant had been sitting continuously for 3 hours).

Subsequently the recorded user logs were evaluated, for the majority of aspects used (10 out of 11) the expected functionality was achieved. However, in the case of the facts responsible for assessing head positions, considerably more warnings were sent by the system than was acceptable. As for manually created input (simulated user postures), this behaviour had not emerged again, examination of the monitoring software developed have begun. Upon inspection of the logs created, it was found that this error was a result of improper calibration: the inaccuracy of the sensor itself was higher than anticipated and the distortions that resulted meant that even when the user evidently adopted a suitable posture, the logged value exceeded the threshold defined as acceptable. Based on this information, the related constraints were adjusted accordingly and this undesirable behaviour of the system was successfully removed. This source of inconvenience, however, highlights the possible sensitivity and dependency of the system, i.e. the reliability of the measurement data received. Still, its significance might be diminished by utilizing methods developed for improving the accuracy and error tolerance of Kinect, such as the ones proposed in [18–20].

In general, it can be said that the system developed and the initial rule set created are capable of providing the required functionality. However, there is still room for improvement as the constraints involved could be fine-tuned and additional or more precise aspects implemented. Moreover, a more complex system capable of adapting to user behaviour could also be created, the main goal of which would be to assist users, without being too rigorous or repetitive, by changing how frequently and in what manner its responses are displayed. This could help to maintain the motivation of workers to break bad habits concerning improper postures whilst seated.

Even though the number of participants in the experiments conducted was sufficient to validate the initial version of the expert system, further investigations with considerably larger user groups are desirable as they would provide a more thorough validation and may yield additional insight into what other aspects the system could provide assistance to users. Additionally, while the recording sessions, on average of 5-6 hours in duration, implemented so far have been useful, by extending these to a few days or even a couple of weeks, more complex behavioural patterns could be identified and analysed. Furthermore, by providing real-time assessment over much longer periods, the effects caused by this intervention could be investigated and subsequently the methods used to assist users improved.

Nevertheless, from the results that have already been logged, it can be clearly seen that the global tendency of users to be seated for longer periods of time than is advisable was also exhibited by the cases investigated. Most participants reported that over such prolonged sessions whilst seated, a need to change position is common, however, usually a better alternative cannot be found. A solution to this problem could be to provide workers with electric height adjustable desks, which may facilitate the ability to switch easily between sitting and standing working postures and customise the height of desks to match the anthropometric features of the individual. Moreover, such furniture could enable the expert system itself to "take action", when needed, to raise the desk when users fail to heed previous warnings.

It should be noted that, when asked, most of the participants considered such a monitoring and assessment system to be useful, however, some, while welcoming the idea of an adjustable desk, expressed concerns about personal security and were reluctant to be "continuously monitored". This reveals that if such a system enters the market, apart from providing the necessary security measures, assuring users that their personal data is safe would also be necessary.

4. Conclusion

In this research a proof-of-concept CLIPS-based rulebase for an expert system was created to facilitate ergonomic evaluation of users in office environments. It has been shown that such a system is capable of thoroughly evaluating user behaviour, thus, implementing interventions in order to decrease the negative effects of prolonged sedentariness.

Acknowledgement

The research was supported by the ÚNKP-18-3 New National Excellence Program. The authors thank the assistance of the Department of Ergonomics and Psychology of the Budapest University of Technology and Economics. The authors also acknowledge the support of the

Széchenyi 2020 programme under the EFOP-3.6.1-16-2016-00015 project.

REFERENCES

[1] Smith, M. J.; Conway, F. T.; Karsh, B. T.: Occupational stress in human computer interaction, *Ind. Health,* 1999 **37**(2), 157–173 DOI: 10.2486/indhealth.37.157

[2] Vignais, N.; Miezal, M.; Bleser, G.; Mura, K.; Gorecky, D.; Marin, F.: Innovative system for real-time ergonomic feedback in industrial manufacturing, *Appl. Ergonomics,* 2013 **44**(4), 566–574 DOI: 10.1016/j.apergo.2012.11.008

[3] Battini, D.; Persona, A.; Sgarbossa, F.: Innovative real-time system to integrate ergonomic evaluations into warehouse design and management, *Computers Ind. Eng.,* 2014 **77**, 1–10 DOI: 10.1016/j.cie.2014.08.018

[4] Zhang, Z.: Microsoft Kinect sensor and its effect, *IEEE Multimedia,* 2012 **19**(2), 4–10 DOI: 10.1109/MMUL.2012.24

[5] Plantard, P.; Shum, H. P. H.; Le Pierres, A. S.; Multon, F.: Validation of an ergonomic assessment method using Kinect data in real workplace conditions, *Appl. Ergonomics,* 2017 **65**, 562–569 DOI: 10.1016/j.apergo.2016.10.015

[6] Cippitelli, E.; Gasparrini, S.; Spinsante, S.; Gambi, E.: Kinect as a tool for gait analysis: Validation of a real-time joint extraction algorithm working in side view, *Sensors,* 2015 **15**(1), 1417–1434 DOI: 10.3390/s150101417

[7] Rivero, L. C.; Rodríguez, R. G.; Pérez, M. D. R.; Mar, C.; Juárez, Z.: Fuzzy logic and RULA method for assessing the risk of working, *Procedia Manufacturing,* 2015 **3**, 4816–4822 DOI: 10.1016/j.promfg.2015.07.591

[8] Kong, Y. K.; Lee, S. Y.; Lee, K. S.; Kim, D. M.: Comparisons of ergonomic evaluation tools (ALLA, RULA, REBA and OWAS) for farm work, *Int. J. Occup. Safety Ergon.,* 2018 **24**(2), 218–223 DOI: 10.1080/10803548.2017.1306960

[9] Golabchi, A.; Han, S.; Fayek, A. R.: A fuzzy logic approach to posture-based ergonomic analysis for field observation and assessment of construction manual operations, *Can. J. Civil Eng.,* 2016 **43**(4), 294–303 DOI: 10.1139/cjce-2015-0143

[10] Berlin, C.; Adams, C.: Production ergonomics: Designing work systems to support optimal human performance, (Ubiquity Press, London, England) 2017 pp.139-160 DOI: 10.5334/bbe

[11] Szakonyi, B.; Lőrincz, T.; Lipovits, Á.; Vassányi, I.: An expert system framework for lifestyle counselling. In: eTELEMED 2018: The tenth international conference on eHealth, telemedicine, and social medicine, 2018 pp. 92-96 ISBN: 978-1-61208-618-7

[12] Riley, G.: CLIPS: a tool for building expert systems, `http:\www.clipsrules.net/?q=AboutCLIPS` [accessed September 20, 2018]

[13] Google Inc.: Protocol Buffers, `http:\developers.google.com/protocol-buffers` [accessed September 20, 2018]

[14] Lueder, R.: A proposed RULA for computer users, In: Proceedings of the ergonomics summer workshop, UC Berkeley Center for Occupational & Environmental Health Continuing Education Program, San Francisco, August 8-9, 1996 **24**, 91–99

[15] Hignett, S.; McAtamney, L.: Rapid Entire Body Assessment (REBA), *Appl. Ergonomics,* 2000 **31**(2), 201–205 DOI: 10.1016/S0003-6870(99)00039-3

[16] Karhu, O.; Kansi, P.; Kuorinka, I.: Correcting working postures in industry: A practical method for analysis, *Appl. Ergonomics,* 1977 **8**(4), 199–201 DOI: 10.1016/0003-6870(77)90164-8

[17] Douwes, M.; Boocock, M.; Coenen, P.; van den Heuvel, S.; Bosch, T.: Predictive validity of the Hand Arm Risk assessment Method (HARM), *Int. J. Ind. Ergonomics,* 2014 **44**(2), 328–334 DOI: 10.1016/j.ergon.2013.09.003

[18] Plantard, P.; Hubert, H. P.; Multon, F.: Filtered pose graph for efficient kinect pose reconstruction, *Multimedia Tools Appl.,* 2017 **76**(3), 4291–4312 DOI: 10.1007/s11042-016-3546-4

[19] Shum, H. P. H.; Ho, E. S. L.; Jiang, Y.; Takagi, S.: Real-time posture reconstruction for Microsoft Kinect, *IEEE Trans. Cybernetics*, 2013 **43**(5), 1357–1369 DOI: 10.1109/TCYB.2013.2275945

[20] Zhou, L.; Liu, Z.; Leung, H.; Shum, H.P.H.: Posture reconstruction using Kinect with a probabilistic model, In: Proceedings of the ACM Symposium on Virtual Reality Software and Technology (VRST), 2014, pp. 117-125 DOI: 10.1145/2671015.2671021