An Impact of the Variable Technological Progress Rate on the Trajectory of Labor Productivity

Monika Bolińska[1], Paweł Dykas[1], Grzegorz Mentel[2], Tomasz Misiak[3]

[1]Department of Mathematical Economics, The Faculty of Management and Social Communication, Jagiellonian University in Cracow, Prof. Stanisława Łojasiewicza 4 St., 30-348 Cracow, e-mail: monika.bolinska@doctoral.uj.edu.pl; pawel.dykas@uj.edu.pl

[2]Department of Quantitative Methods, Faculty of Management, Rzeszow University of Technology, Powstancow Warszawy 12 St., 35-959 Rzeszow, e-mail: gmentel@prz.edu.pl

[3]Department of Economics, Faculty of Management, Rzeszow University of Technology, Powstancow Warszawy 12 St., 35-959 Rzeszow, e-mail: tmisiak@prz.edu.pl

Abstract: The goal of this paper is to develop the neo-classical Solow growth model, in which, the authors repeal an assumption of a constant rate of growth of technological progress. Herein, the authors assume an alternative trajectory of an increase in scientific and technical knowledge A(t), and on that basis, they accept the following assumptions. First, the rate of growth of technological progress is not constant, but changes over time. Second, the path of the growth of the scientific and technological knowledge tends toward a certain level in the long term, which can be equated with the equivalent of the technological boundary. Such a modification of the assumptions regarding the technological progress rates, allows leading growth paths for both capital and product per unit of effective labor. Next, based on the solution of the presented growth model, the authors calibrated the parameters and carried out numerical simulations. Numerical simulations, conducted for the Polish economy in a 100-year horizon, allowed an inclusion of scenarios regarding both the rate of technological progress and investment rates. In the simulations, two variants of the annualized rates of technological progress were adopted (optimistic, g=1.7% and realistic, g=1.5%). The adopted levels of technological progress rates were used to determine the horizontal asymptote for the scientific and technological knowledge A(t) that will be shaped in accordance with geometric progress. In the considered variants the following investment rates were adopted: 15, 20 and 25%. This allowed determining the trajectories of labor productivity growth in the Polish economy taking into account different combinations of changes in the rates of technological progress and investment rates.

Keywords: trajectories of labor productivity; variable rates of technological progress; gamma function; numerical simulations

1 Introduction

In the theory of economics technological progress is considered as one of the most important factors determining economic growth. In growth models technological progress is perceived differently and there is a large variety of factors that generate progress, which undoubtedly affects its exogenous or endogenous character. Apart from the factors determining technological progress, in this article the authors attempt to develop the neo-classical growth Solow model. They repeal the assumption of a constant growth rate of technological progress. In the paper the authors assume an alternative trajectory of an increase in scientific and technological knowledge, and on its basis of they assume the following assumptions. Firstly, the growth rate of technological progress is not constant but changes over time. Secondly, the growth path of scientific and technological knowledge tends to a certain level in the long term, which can be equated with the equivalent of the technological boundary.

The structure of this paper is as follows. The first section is an introduction. The second part, a review of the literature regarding the inclusion of technological progress in selected models of economic growth and selected generalizations of the neo-classical Solow model were presented. The third part contains an analytical solution of the model that takes into account alternative assumptions regarding the development path of the scientific and technological knowledge. In the fourth part, the parameters of the presented model are calibrated and numerical simulations of labor productivity growth paths in the considered variants are presented. Part five is a summary of the considerations and the more important conclusions.

2 Review of the Literature

Economic growth is a multidimensional and long-term process leading to an increase of the production potential of a given economy. The multidimensional character of economic growth is primarily the result of many factors determining this process, and the basic ones include: capital, labor, technological progress, institutional factors or social capital [25, 27, 30]. Multidimensionality also results from an analysis of various aspects of particular factors that determine the processes of economic growth [26]. Capital in growth models is considered as material, human or social capital. Technological progress as an exogenous or endogenous factor, embodied or unmasked in a man, etc. A great contribution to

the development of the theory of growth is attributed to the Solow [23] model which became the foundation for later growth models. The Solow model was based on the power production function of the Cobb-Douglas type and the equation of capital accumulation which determines its growth as investments in this capital reduced by its depreciated value. This model also assumed a constant income from the scale of the production process, a constant rate of technological progress, or a constant rate of increase in the number of employees [23]. The basic advantages of the Solow model include the form of the equation that describes the accumulation of tangible capital, an uncomplicated analytical form of the model solution, which, with the power production function, is easily empirically verifiable by determining the growth paths of technological labor or labor productivity. The main weaknesses of the model include, first of all, an adoption of physical capital as the only factor of production, thus omitting, for example, human capital, adopting permanent economies of scale and treating technological progress or the number of employees as exogenous variables (for more on the disadvantages and critiques of the Solow model [18] [19]. Criticism of the exogenous nature of technological progress in the Solow model made that many researchers attempted to endogenize a technological progress in growth models. One of the first researchers who presented the concept of endogenization of technological progress was Conlisk. In his research he accepted the assumption that economy and its growth are a closed system of dependencies between investment rates in human and material capital and technological progress [6]. Conlisk's assumptions were developed as part of a new growth theory in which the so-called Solow residual was understood as the result of expenditure in human capital, the product of which are new production techniques determining economic growth. Technological progress as a product of human capital was endogenized. As part of the new growth theory two main trends of research can be distinguished. In the first approach an increase in technological progress is the effect of the accumulation of knowledge. In the endogenization models of knowledge accumulation, the research sector (R&D sector) is introduced. Its purpose is to generate new knowledge, and thanks to it, a larger product can be obtained with a given capital and labor resource. Such an approach to the endogenization of technological progress occurred in works including: [3] [20] [11] [1] [28] and empirical studies in the works of, among others [16] [29] [14]. The second trend assumes that the accumulation of capital is of fundamental importance for growth, while capital is recognized more broadly, including human capital as the causative factor of technical progress. The models created under the second trend include, among others works: Romer [21], Barro [2], Mankiw, Romer, Weil [17] or Kramer, Thompson [15].

Theories of endogenous growth, although they justified the issue of generating technological progress, proved to be incapable of explaining the key empirical regularities in the processes of growth and development in various regions of the world during the last two millennia. In general, exogenous and endogenous theories are classified as non-uniform theories of economic growth. Currently, one

of the most interesting and advanced theories is Unified Growth Theory – Oded Galor [9]. With regard to this theory, it can be said, that it belongs to neoclassical theories of endogenous growth. One of its main assumptions is an interaction between investments in human capital and technological progress which evoke the following spiral. Human capital generated faster technological progress, which in turn increased the demand for human capital, leading to increasing investments in children quality, which ultimately led to a decrease in demographic growth and, as a result, to a decline in the population. An increase in demand for human capital caused by technological progress eventually resulted in an increase in the quality of children's education at the expense of the number of children they have. Therefore, human capital also plays an important role here. It creates a faster technological progress but also determines the demand for human capital, which results in the growth of human capital while leading in highly developed economies to a decrease in the birth rate and population, and this becomes a new stimulus for a further technological progress. The technology growth rate in the Galor [9] model is defined as a function of the quality of education and a sufficiently large active population. However, it is not constant as in the Solow model but is subject to change. In addition, the technological boundary is possible, and in the Solow model technological progress tends to infinity.

Most growth models were based on certain weaknesses of the Solow model, such as the assumption of the exogenous nature of technological progress or the adoption of physical capital as the only factor of production, excluding human capital. Some researchers focused their attention on developing the Solow model by repealing rather unrealistic assumptions. An example of this may be the work where the assumption that there is a constant rate of employment growth (constituting a constant percentage of the exponentially growing population) exists. It means that the number of people working in the economy in the long run increases to infinity, which was dictated by the existence of a positive rate of demographic growth. Nowadays, demographic processes taking place in many developed economies undermine the accuracy of the above assumptions. The studies where alternative assumptions about the paths of employment growth are considered include the works: Guerrini [10] [12], Biancia, Guerrini [4], Sinnathurai [22], Sika, Vidová [24] or Dykas, Misiak, Mentel [7].

In this paper the authors attempt to develop the neo-classical Solow economic growth model and they repeal the assumption of a constant growth rate of technological progress. In the study the authors assume an alternative trajectory of an increase in scientific and technological knowledge, and on its basis they introduce the following assumptions. Firstly, the growth rate of a technological progress is not constant but changes over time. Secondly, the growth path of the scientific and technological knowledge tends to a certain level in the long term, which can be equated with the equivalent of the technological boundary.

3 Model

The model of economic growth discussed in this paper is based on the following assumptions:

1) The production process is shaped according to the Cobb-Douglas [5] power production function according to the formula (see also: Tokarski [27; 28])[1]:

$$Y(t) = (K(t))^{\alpha}(E(t))^{1-\alpha} \tag{1}$$

where Y is the stream of produced product, K and E are (respectively) capital expenditure and so-called the units of effective work, $\alpha\epsilon(0;1)$ is the product's flexibility in terms of capital expenditure, and $1\text{-}\alpha\epsilon(0;1)$ is the flexibility of the product created in the economy relatively to the units of effective work.

2) The accumulation of capital, as it is in the original Solow [23] model, is described by the following differential equation:

$$\dot{K}(t) = sY(t) - \delta K(t) \tag{2}$$

where $\delta\epsilon(0;1)$, $s\epsilon(0;1)$ mean (respectively) the investment rate and the rate of capital depreciation.

3) The amount of scientific and technological knowledge at the t moment is shaped according to the following growth path:

$$A(t) = A_0 - \theta e^{-\lambda t} \tag{3}$$

where: $A_0\text{-}\Theta>0$

The growth path of the scientific and technological knowledge described by equation (3) is characterized by the fact that in the infinite time horizon ($t\rightarrow\infty$) this resource tends to the level of A_0. However, in the period t=0, the scientific and technological knowledge in the discussed model of economic growth is at the level $A_0\text{-}\Theta>0$. Moreover, by differentiating the relation (3) with respect to time, we obtain: $\dot{A}(t) = \theta\lambda e^{-\lambda t} > 0$. The above dependences show that the growth path described by (3) grows asymptotically from level $A_0\text{-}\Theta$ to the value A.

4) Effective work units $E(t)$ are defined as a set of scientific and technological knowledge $A(t)$ weighted by the number of employed $L(t)$. When determining the growth rate of units of effective work $\dot{E}(t)/E(t)$, it turns out that they grow at a growth rate equal to $n+g(t)$, where $n> 0$ is the rate of growth in the number

[1] All subsequent macroeconomic variables are assumed to be differentiable functions of time $t\geq0$. The record $x(t)$ will mean the value of the variable x at the moment t, and $x(t)=dx/dt$ - the derivative of the variable x after the time t, i.e. (economically speaking) an increase in the value of this variable at the moment t.

of the employed in this economy, and $g(t) = \dot{A}(t)/A(t)$ is the growth rate of the scientific and technological knowledge referred to in the literature as technological progress rate. Hence and from dependence (3), it follows that the rate of technological progress in the discussed economic growth model is the following:

$$g(t) = \frac{\theta\lambda e^{-\lambda t}}{A_0 - \theta e^{-\lambda t}} \tag{4}$$

From the above considerations, it follows that the growth rate of the effective work units is in line with:

$$\frac{\dot{E}(t)}{E(t)} = \frac{\theta\lambda e^{-\lambda t}}{A_0 - \theta e^{-\lambda t}} + n \tag{5}$$

5) Assuming that $y=Y/L$ and $k=K/L$ are (respectively) labor productivity and technical development, then the following relationships will be satisfied:

$$Y(t) = L_0 e^{nt} y(t) \tag{6}$$

$$K(t) = L_0 e^{nt} k(t) \tag{7}$$

6) Moreover, assuming that $y_E=Y/E$ and $k_E=K/E$ $yE=Y$ are (respectively) the stream of product produced and the capital stock per unit of effective work and based on equation (6) - (7) we get:

$$y(t) = \left(A - \theta e^{-\lambda t}\right) y_E(t) \tag{8}$$

$$k(t) = \left(A - \theta e^{-\lambda t}\right) k_E(t) \tag{9}$$

From the production function (1) one can go to the production function in an intense form by dividing its sides by units of effective work $E>0$, which based on assumption 6) gives:

$$y_E(t) = \left(k_E(t)\right)^{\alpha} \tag{10}$$

The relation (10) describes the relationship between the capital expenditure per unit of effective work (k_E) and the production volume per unit of work (y_E).

Differentiating capital for the unit of effective work($k_E=K/E$) after the t time we obtain:

$$\dot{k}_E(t) = \frac{\dot{K}(t)E(t) - K(t)\dot{E}(t)}{\left(E(t)\right)^2} = \frac{\dot{K}(t)}{E(t)} - \frac{\dot{E}(t)}{E(t)} k_E(t)$$

which together with (2)-(5) gives us:

$$\dot{k}_E(t) = s y_E(t) - \mu(t) k_E(t) \tag{11}$$

and $\mu(t)=\delta+g(t)+n>0$ means the rate of capital loss per unit of effective work. The differential equation (11) is equivalent to the Solow movement equation [22] in the economic growth model discussed here.

Taking into account the production function in the intense form (10) and the relation (11), one can obtain the following differential equation:

$$\dot{k}_E(t)=s(k_E(t))^{\alpha}-\mu(t)k_E(t) \tag{12}$$

The equation (12) for each $t\geq 0$ has a trivial solution ($k_E(t)=0$) and a family of non-trivial integrals[2].

The equation (12) for $k_E>0$ can be presented as:

$$(k_E(t))^{-\alpha}\dot{k}_E(t)=s-\mu(t)(k_E(t))^{1-\alpha} \tag{13}$$

By making the Bernoullei substitution:

$$z(t)=(k_E(t))^{1-\alpha} \tag{14}$$

we get the following transformation of the equation (13):

$$\frac{\dot{z}(t)}{1-\alpha}=s-\mu(t)z(t)$$

which can be transformed into the relation:

$$\dot{z}(t)=(1-\alpha)s-(1-\alpha)\mu(t)z(t) \tag{15}$$

Considering the homogeneous equation from the relation (15) we get:

$$\dot{z}(t)=-(1-\alpha)\mu(t)z(t) \tag{16}$$

the solution of equation (16) is given by the formula:

$$z(t)=C(t)e^{-(1-\alpha)(\delta+n)t}(A(t))^{\alpha-1} \tag{17}$$

where the factor $C(t)$ is the integral integration constant. Differentiating the equation (17) with respect to time and taking into account the relation (14) we get:

[2] The trivial integral (as uninteresting from both mathematical and economic point of view) will be further ignored. Non-trivial integral of this equation will determine the time path (or path of growth) of capital for the unit of effective work.

$$(1-\alpha)s-(1-\alpha)\left(\delta+n+\frac{\dot{A}(t)}{A(t)}\right)C(t)e^{-(1-\alpha)(\delta+n)t}\left(A(t)\right)^{\alpha-1}=$$

$$\dot{C}(t)e^{-(1-\alpha)(\delta+n)t}\left(A(t)\right)^{\alpha-1}+C(t)(\alpha-1)(\delta+n)e^{-(1-\alpha)(\delta+n)t}\left(A(t)\right)^{\alpha-1}+$$

$$+C(t)e^{-(1-\alpha)(\delta+n)t}(\alpha-1)\left(A(t)\right)^{\alpha-2}\dot{A}(t)$$

Thus, and from the relations (3) - (5) we get:

$$C(t)=(1-\alpha)s\int\exp\left((1-\alpha)(\delta+n)t\right)\left(A_0-\theta\exp(-\lambda t)\right)^{1-\alpha}dt$$

Selecting t so that the condition is met[3]:

$$\lambda t=\tau\lambda+W\left(-\theta\lambda\exp(-\lambda\tau)\right)$$

we get $C(t)$ approximated to the form:

$$C(t)=(1-\alpha)s\int\exp\left(t(1-\alpha)(n+\delta)\right)t^{1-\alpha}dt$$

In addition (for $\omega=2-\alpha$) by differentiating $\Gamma(\omega,(1-\omega)(n+\delta)t)$ we get[4]:

$$\frac{d\left(\Gamma\left(\omega,(1-\omega)(n+\delta)t\right)\right)}{dt}=e^{(\omega-1)(n+\delta)t}t^{\omega-1}(1-\omega)(n+\delta)^{\omega}$$

From here we finally get:

$$C(t)=s(\omega-1)^2\left((1-\beta)(\delta+n)\right)^{-\omega}\Gamma\left(\omega,(1-\omega)(n+\delta)t\right)+C$$

where $C>0$.

Hence, and from the relations (14) and (17), capital for the unit of effective work can be written as:

$$k_E(t)=\left(e^{(1-\omega)(\delta+n)t}\left(A(t)\right)^{\omega-1}s(\omega-1)^2\left((1-\beta)(\delta+n)\right)^{-\omega}\left(\Gamma\left(\omega,(1-\omega)(n+\delta)t\right)+C\right)\right)^{\frac{1}{\omega-1}} \qquad (18)$$

Assuming that for the discussed problem the Cauchy boundary condition of takes the form $k_E(0)=k_{E0}\geq 0$, the constant $C>0$ can be written as:

$$C=\frac{k_{E0}^{\beta-1}-(\tau-\theta)^{1-\beta}sA_0^{1-\alpha}(\beta-1)\left((1-\beta)(\delta+n)\right)^{1-\beta}\Gamma(\beta,0)}{(\tau-\theta)^{1-\beta}}$$

[3] The function $W(z)$ denotes the special function of W-Lambert, i.e. the function which for the complex number z meets the relation: $z=W(z)\exp(W(z))$.

[4] The function $\Gamma(z)$ is a special gamma function, i.e. a function defined for any complex number as: $\Gamma(z)=\frac{1}{z}\prod_{n=1}^{\infty}\frac{\left(1+\frac{1}{n}\right)^z}{1+\frac{z}{n}}$.

Thus, the integral of equation (13) is written as:

$$k_E(t) = \Big[sA_0^{1-\alpha}(\beta-1)\big((1-\beta)(\delta+n)\big)^{1-\beta}\Gamma(\beta,(1-\beta)(n+\delta)t) + \frac{k_{E0}^{\beta-1} - (\tau-\theta)^{1-\beta}sA_0^{1-\alpha}(\beta-1)\big((1-\beta)(\delta+n)\big)^{1-\beta}\Gamma(\beta,0)}{(\tau-\theta)^{1-\beta}} \Big(e^{t(1-\beta\delta)}\big(\tau-\theta e^{-\lambda t}\big)^{1-\beta}\Big)\Big]^{\frac{1}{1-\alpha}} \quad (19)$$

In addition, the product for the unit of effective work ($y_E(t)$) is described by the equation:

$$y_E(t) = \Big[sA_0^{1-\alpha}(\beta-1)\big((1-\beta)(\delta+n)\big)^{1-\beta}\Gamma\big(\beta,(1-\beta)(n+\delta)t\big) + \frac{k_{E0}^{\beta-1} - (\tau-\theta)^{1-\beta}sA_0^{1-\alpha}(\beta-1)\big((1-\beta)(\delta+n)\big)^{1-\beta}\Gamma(\beta,0)}{(\tau-\theta)^{1-\beta}} \cdot \Big(e^{t(1-\beta\delta)}\big(\tau-\theta e^{-\lambda t}\big)^{1-\beta}\Big)\Big]^{\frac{\alpha}{1-\alpha}} \quad (20)$$

4 Calibration of Model Parameters and Numerical Simulations

Numerical simulations of the theoretical model presented in point 3 were conducted in two stages in one-hundred-year periods. In the first stage, the values of parameters of equations describing the shaping of the scientific and technological knowledge and the path of labor productivity growth were calibrated. In the second stage, numerical simulations of the above-mentioned paths were performed based on different scenarios regarding the investment rates and the technological progress rate. For the function describing the growth path of the scientific and technological knowledge base, the following conjunction was adopted:

$$\begin{aligned} &A_0\text{-} \\ &\Theta = A_{2000} \\ &\wedge \\ &A_0 = A_{2100} \end{aligned} \quad (21)$$

where: A_{2000} and A_{2100} are (respectively) resources of scientific and technological knowledge in 2000 and 2100. In the study, the level of scientific and technological knowledge $A(t)$ in 2000 was normalized to unity, and the level for 2100 was determined based on the average annual technological progress rate assuming that $A(t)$ will be shaped in accordance with geometric progress. To this end, the

production function (1) was used, assuming a constant rate of increase in scientific and technological knowledge $\frac{\dot{A}(t)}{A(t)} = g > 0$, thus, we get:

$$Y(t) = \tilde{A}e^{gt}\left(K(t)\right)^{\alpha}\left(L(t)\right)^{1-\alpha} \tag{22}$$

Where $\tilde{A} > 0$ means the level of scientific and technological knowledge in the period t=1.

By logarithmizing equation (22) we obtain:

$$\ln\left(Y(t)\right) = \ln\left(\tilde{A}\right) + gt + \alpha\ln\left(K(t)\right) + \left(1-\alpha\right)\ln\left(L(t)\right) \tag{23}$$

When subtracting, in equation (23), the items $ln(L(t))$ one can go to dependencies:

$$\ln\left(y(t)\right) = \ln\left(\tilde{A}\right) + gt + \alpha\ln\left(k(t)\right) \tag{24}$$

The equation (24) describes the log-linear relationship between labor productivity y and the technological development k and the rate of technological progress. In the next stage, based on panel data for Polish provinces taken from the Central Statistical Office for the years 2000-2015, the parameters of the following equation were estimated:

$$\ln\left(y_{it}\right) = \alpha_0 + \alpha_1 t + \alpha_2\ln\left(k_{it}\right) \tag{25}$$

where y_{it} is labor productivity in the i-th of the provinces (i=1,2, ..., 16) in year t (t= 2000, 2001, ..., 2015); k_{it} is a technological development in the i-th province in year t; α_0 this is the logarithm of the total productivity of factors of production; α_1 parameter determining technological progress; α_2 flexibility of labor productivity in relation to technological development.

Table 1
Estimated parameters of the equation (25)

Explanatory variable	Values of estimated parameters
Constant	-32.0685 (0.0000)
$\ln(k_{it})$	0.2997 (0.0000)
T	0.0174 (0.0000)
R^2 Corr. R^2	0.8171 0.8151

Source: personal study

The level of significance was given in brackets.

The estimation of the parameters of the equation (25) was made on the basis of the instrumental variables method using the Generalized Method of Moments. The estimates presented in table 1 show that the average flexibility of labor productivity in relation to technological development in Polish provinces in the years 2000-2015 was about 0.30 and the value of such flexibility was accepted for further numerical simulations (see also [8]). In addition, the average annual technological progress rate for this period was approx. 1.7%. For this reason, in numerical simulations two variants were adopted regarding the formation of the A(t) value. The first variant (called optimistic) assumes such level of A(t) in 2100 that would be determined by the average annual technological progress rate consistent with the estimates, while the second variant (realistic) assumes an average annual growth of 1.5%[5].

Figure 1 presents the growth paths of scientific and technological knowledge with the assumed variants. Assuming the first variant, where the level $A(t)$ in 2100 will be implied by an average annual technological progress equal to 1.7%. Assuming such a scenario of shaping the scientific and technological knowledge, its value in 2100 in relation to 2000 will increase by 4.8 times.

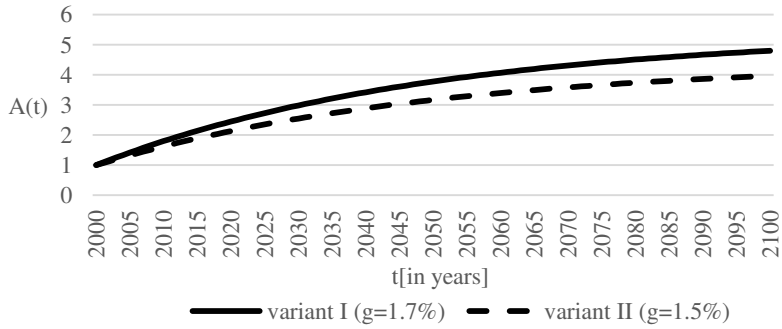

Figure 1
Trajectories of the scientific and technological knowledge in the adopted variants

Source: own study

[5] The Polish economy has undergone a system transformation from a centrally planned economy to a market economy. Almost 50 years of existence of a centrally planned economy in Poland generated significant technological delays (a technological gap) in relation to the economies of Western Europe. In the period 2000-2015, which was adopted to estimate the rate of technological progress, the Polish economy on the principle of technological convergence continued to reduce the technological gap. Thus, the estimated rate of technological progress may be overstated as it contains the effect of technological convergence. For this reason, in the realistic variants, the average annual rate of technological progress at the level of 1.5% was adopted.

While accepting variant II an increase in the amount of scientific and technological knowledge in 2100 will be fourfold in comparison with 2000.

In addition, in each of the variants considered, three scenarios regarding the investment rates were adopted. In the years 2000-2015, the average investment rate for the Polish economy was at the level of 19.8%, and based on this average the authors assumed that this rate in the discussed time horizon would be equal to 20% assuming that it may deviate by 5 points rates. Bearing in mind the above, the variants regarding the investment rate were: 15%, 20%, 25%. The growth rate of the number of employees was assumed at 1%. In addition, based on the capital accumulation equation (2), the rate of depreciation of capital for the Polish economy was estimated. The equation describing the rate of depreciation of capital in discrete time is as follows:

$$\delta = \frac{sK_t^{\alpha} - \Delta K_t}{K_t} \tag{26}$$

Equation (26) assumes flexibility of labor productivity in relation to capital-labor ratio at the level of 30%, and the investment rate at 20%. Based on statistical data taken from the Central Statistical Office, regarding the development of physical capital in the Polish economy for the years 2000-2015, the rate of depreciation of capital at the level of 8.5% was estimated.

Table 2

Numerical simulations of labor productivity in various variants of the rate of technological progress and investment rates

Simulation period (in years)	Variant I (g=1.7%)			Variant II (g=1.5%)		
	s=0.15	s=0.2	s=0.25	s=0.15	s=0.2	s=0.25
2000	1	1	1	1	1	1
2020	3.611	4.037	4.410	3.014	3.366	3.674
2040	6.333	7.152	7.861	5.0181	5.665	6.225
2060	8.498	9.609	10.571	6.589	7.451	8.196
2080	10.071	11.391	12.534	7.725	8.738	9.614
2100	11.172	12.638	13.906	8.518	9.636	10.603

Source: Own study

Table 2 presents numerical simulations for a 100-year time horizon for the Polish economy. The following conclusions of the economic character can be drawn from the results of numerical simulations regarding the labor productivity growth paths.

When assuming the first variant (see Fig. 2) regarding the rate of technological progress, and assuming that the economy will be characterized by a relatively low investment rate of 15%, labor productivity in the discussed time horizon will increase more than eleven times. In the same scenario concerning technological progress, but assuming that investment rates will be at 20%, labor productivity in the Polish economy will increase by about 12.5 times as compared to 2000. With the first option, the highest increase in labor productivity would occur at the rate investment equal to 25% and this increase would be almost fourteen.

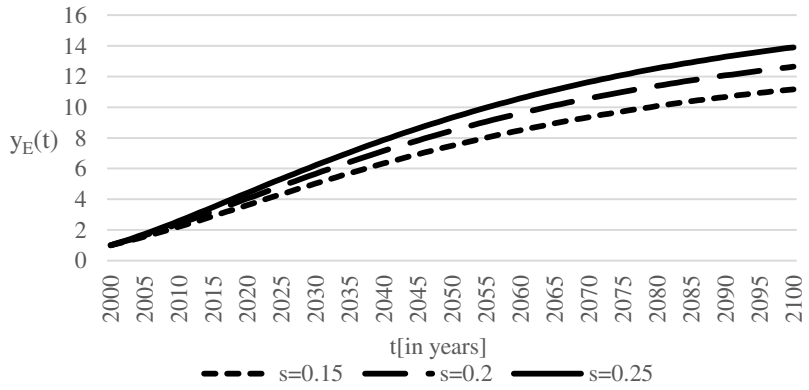

Figure 2
Labor productivity growth paths in various scenarios regarding investment rates and technological progress rate g=1.7%

Source: Own study

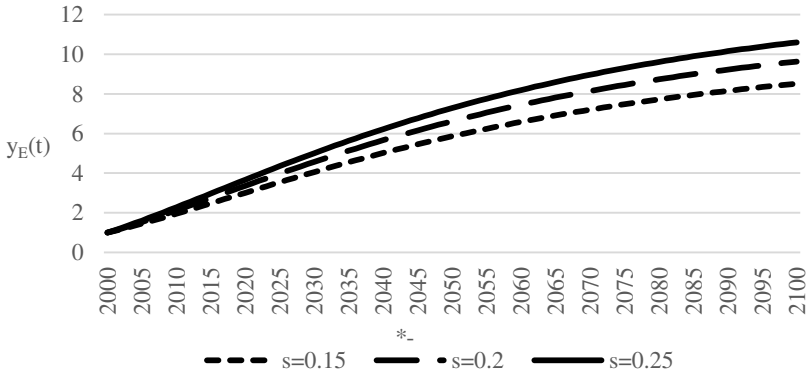

Figure 3
Labor productivity growth paths for various scenarios regarding investment rates and technological progress rate g=1.5%

Source: Own study

When assuming that the level of the scientific and technological knowledge in the year 2100 will increase by 4.8 times - the second variant and also assuming that the investment rate in the discussed time horizon will be 15%, then the labor productivity will increase by about 8.5 times. However, with the same variant, assuming an investment rate of 20%, the product per employee in the economy up to 2050 will increase by about 9.6 times. On the other hand, Figure 3 shows that with the second variant regarding the rate of technological progress, with an investment rate of 25%, labor productivity in the years 2000-2050 will increase by around 10.6 times.

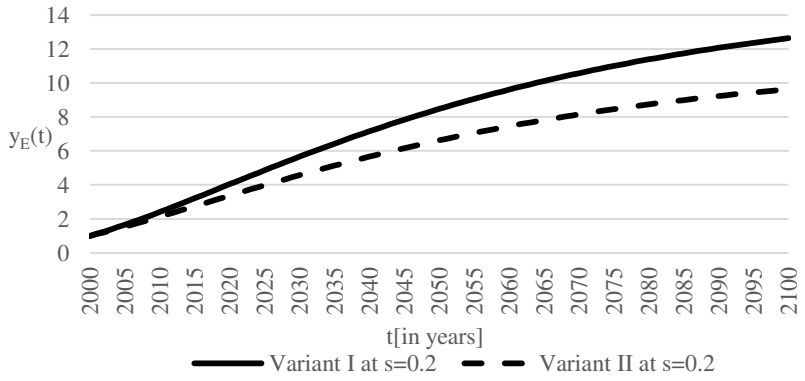

Figure 4
Compilation of labor productivity growth paths for different variants concerning the rate of technological progress and the investment rate s=20%

Source: Own study

When comparing two options regarding the development of the scientific and technological knowledge in the horizon 2000-2100, it can be noticed that, for example, in the case of investment rates of 20% (see Fig. 4), labor productivity in 2100 will be approx. 30% higher than for the second variant.

Conclusions

The model of economic growth presented in the paper is a modification of the neo-classical Solow-Swan model (1956). In the model under consideration, the assumption of a constant rate of technological progress was abolished, thus assuming the path of growth of the scientific and technological knowledge resources changing exponentially to a permanent asymptote. This modification allowed taking into account, herein, certain scenarios concerning the shaping of the technological progress rate, thus obtaining various paths of growth of the scientific and technological knowledge. The study adopted two scenarios, regarding the technological progress rate, the first optimistic variant with a progress rate equal to 1.7% and (the second) realistic one for the rate equal to

1.5%. Numerical simulations show that the scientific and technological knowledge in the Polish economy, 2000-2100 will increase about 4.8 times in the implementation of the optimistic variant, while the realistic variant has a lower growth rate of 400% in the year 2100 compared to the year 2000.

For each of the variants regarding the development of the technological progress rate, three scenarios for the investment rate were selected. The following levels: 15%, 20%, and 25% were assumed. The numerical simulations carried out for the investment rate of 15% allow noticing that depending on the adopted scenario, an 11-fold increase in labor productivity (for the optimistic variant) and an 8.5-fold increase for the realistic variant is possible. With an investment rate of 20%, an increase in labor productivity in the discussed time horizon was 12.6 times (in the optimistic variant) or 9.6 times (in the realistic variant). The highest increase in labor productivity was recorded at the investment rate of 25% and for the optimistic variant, it was 13.9 times, while for the realistic variant, it was 10.6 times.

References

[1] Aghion, P., Howitt, P.: A Model of Growth Through Creative Destruction, Econometrica, Vol. 60, March 1992

[2] Barro, R.: Economic Growth in a Cross Section Countries, Working Paper No. 3120, NBER, September 1989

[3] Baumol, W. J.: Entrepreneurship. Productive, Unproductive and Destructive, Journal of Political Economy, Vol. 98, 1990

[4] Bianca, C., Guerrini, L.: Existence of Limit Cycles in the Solow Model with Delayed-Logistic Population Growth, Scientific World Journal, 2014

[5] Cobb, C. W., Douglas, P. H.: A Theory of Production, American Economic Review, No. 18, 1928

[6] Conlisk, J.: A modified neoclassical growth model with endogenous technological change, The Southern Economic Journal, October 1967

[7] Dykas, P., Mentel, G., Misiak, T.: The Neoclassical Model of Economic Growth and Its Ability to Account for Demographic Forecast, Transformations in Business & Economics, Vol. 17, No 2B(44B), 2018, pp. 684-700

[8] Filipowicz, K., Misiak, T., Tokarski, T.: Bipolar growth model with investment flows, Economics and Business Review, Vol. 2(16) No. 3, 2016

[9] Galor, O.: Unified Growth Theory, Princeton University Press, Princeton & Oxford, 2011

[10] Grossman, G. M., Helpman, E.: Innovation and Growth in the Global Economy, MIT Press, Cambridge 1991

[11] Guerrini, L.: A Closed Form Solution to the Ramsey Model with Logistic Population Growth, Economic Modeling, 27, 2010a, pp. 1178-1182

[12] Guerrini, L.: Logistic Population Change and the Mankiw-Romer-Weil Model, Applied Sciences, 12, 2010, pp. 96-101

[13] Guerrini, L.: The Solow-Swan Model with the Bounded Population Growth Rate, Journal of Mathematical Economics, 42, 2006, pp. 14-21

[14] Jones, Ch.: Time Series of Endogenous Growth Models, Quarterly Journal of Economics, Vol. 110, May 1995

[15] Kremer, M., Thomson, J.: Young Workers, Old Workers, and Convergence, NBER Working Papers, No. 4827, August 1994

[16] Kremer, M.: Population Growth and Technological Change. One Million B.C. to 1990, Quarterly Journal of Economics, Vol. 108, August 1993

[17] Mankiw, N. G., Romer, D., Wei, l D. N.: A Contribution to the Empirics of Economic Growth, Quarterly Journal of Economics, May 1992

[18] McCombie, J. S. L.: The Solow Residual, Technological Change and Aggregate Production Functions, Journal of Post Keynesian Economics, 23 (2), 2000, pp. 267-297

[19] McCombie, J. S. L.: What Does the Aggregate Production Function Tell Us? Second Thoughts on Solow's, Second Thoughts on Growth Theory, Journal of Post Keynesian Economics, 23 (4), 2001, pp. 589-615

[20] Romer, P. M.: Endogenous Technological Growth, Journal of Political Economy, Vol. 98, No. 5, 1990

[21] Romer, P. M.: Increasing Returns and Long-Run Growth, Journal of Political Economy, Vol. 94, No. 86, 1986

[22] Sinnathurai, V.: An Empirical Study on the Nexus of Poverty, GDP Growth, Dependency Ratio and Employment in Developing Countries, Journal of Competitiveness, Vol. 5, Issue 2, June 2013, pp. 67-82

[23] Solow, R. M.: A Contribution to the Theory of Economic Growth, Quarterly Journal of Economics, February 1956. s

[24] Sika, P., Vidová, J.: Interrelationship of migration and housing in Slovakia. Journal of International Studies, 10(3), 91-104, 2017

[25] Simionescu, M., Lazányi, K., Sopková, G., Dobeš, K., Balcerzak, A. P.: Determinants of Economic Growth in V4 Countries and Romania. Journal of Competitiveness, Vol. 9, Issue 1, pp. 103-116, 2017

[26] Sinicakova, M., Gavurova, B.: Single Monetary Policy versus Macroeconomic Fundamentals in Slovakia. Ekonomicky casopis, 65(2): 158-172, 2017

[27] Soltes, V., Gavurova, B.: Modification of Performance Measurement System in the Intentions of Globalization Trends. Polish Journal of Management Studies, 11(2), 160-170, 2015

[28] Strielkowski, W., Tumanyan, Y., Kalyugina, S.: Labour Market Inclusion of International Protection Applicants and Beneficiaries, Economics and Sociology, Vol. 9, No 2, pp. 293-302, 2016

[29] Szilágyi, G. A.: Exploration Knowledge Sharing Networks Using Social Network Analysis Methods. Economics and Sociology, 10(3), 179-191, 2017

[30] Tkacova, A., Gavurova, B., Behun, M.: The Composite Leading Indicator for German Business Cycle. Journal of Competitiveness, Vol. 9, Issue 4, pp. 114-133, 2017

[31] Tokarski, T.: Ekonomia matematyczna. Modele makroekonomiczne, Polskie Wydawnictwo Ekonomiczne, Warszawa 2011

[32] Tokarski, T.: Matematyczne modele wzrostu gospodarczego (ujęcie neoklasyczne), Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków 2009

Reducing the Computational Requirements in the Mamdani-type Fuzzy Control

József Dombi[1], Edit Tóth-Laufer[2]

[1]Faculty of Science and Informatics, University of Szeged, Aradi vértanúk 1, H-6720 Szeged, Hungary, dombi@inf.u-szeged.hu

[2]Bánki Donát Faculty of Mechanical and Safety Engineering, Óbuda University, Népszínház 8, H-1081 Budapest, Hungary, laufer.edit@bgk.uni-obuda.hu

Abstract: Fuzzy-based control systems are popular in practical applications where imprecision, subjectivity, and uncertainty can arise in the data and in the evaluation process, and it needs to be addressed. One possible solution is the fuzzy approach. However, computational requirements of these models can be extremely high, mainly defuzzification part of the Mamdani-type control. Whatever great advantages the Mamdani model has, it is closer to the human way of thinking compared to the Sugeno model, which is another popular controller. In the case where a short reaction time is required, the computational needs should be reduced. Here, we propose modified Mamdani models for this purpose. This new model is much faster while all the advantageous properties of the original Mamdani model are retained, and in some ways, it is better.

Keywords: fuzzy control; Mamdani control; arithmetic-based control

1 Introduction

Fuzzy-based control models are frequently used in engineering tasks and in medical related applications [1]. Because of this, they can be built up in a user-friendly way, i.e. we can apply linguistic terms for the hard-to-quantify parameters. The Mamdani control is closer to the human way of thinking, but it has high computational requirements. For this reason, the application of the model in real-time and adaptive systems is limited. Computational requirements of the Takagi-Sugeno system is much lower. For this reason, it is very popular in optimization and adaptive tasks, but its limitation compared to the Mamdani control that intuition can be less built into the system. For the latter reason, it was not examined in detail in this paper, only a brief comparison is performed.

In the literature, there are several different techniques available for reduces the computational requirements in Mamdani-type control [2], such as the HOSVD-

based model, which eliminates the redundancy from the system [3]; genetic algorithms are used for generalization of the antecedents [4]; and, to minimize the number of rules and membership functions of the fuzzy system fuzzy subtractive clustering [5] or the combination of the inputs or antecedent sets [6] are the most popular approaches. Using a hierarchical structure of the model instead of a single-stage system also decreases the computational needs because the number of the rules increases exponentially with the number of the input parameters. In hierarchical systems, this increase is only linear because the problem is broken down into sub-problems. In this way, there are fewer input parameters at the different stages, leading to sub-systems that operate with fewer rules [7].

In this study, two significant modifications of the conventional Mamdani control system are introduced. One of them is the Mamdani-like structure with a discretized output, where the order of the evaluation steps is swapped, while the second is an arithmetic-based model, which is based on a linear combination of the membership function components. The aim of the study is to find an algorithm that can retain the advantages of the conventional control model, yet significantly reduce the computational requirements. The authors examined the proposed methods in a hierarchical patient surveillance system, where physiological parameters are monitored in real-time.

2 Mamdani-Type Control System

In the Mamdani system natural language rules are used, where the rules are given in the following form: *IF condition THEN consequence*. Let the input parameters be $x_1, x_2, \ldots, x_n$ from the universe of discourse $X_1, X_2, \ldots, X_n$, respectively. In this case, the output parameters $y \epsilon Y$ can be calculated by using the following type of rules:

$$IF\ x_1\ is\ A_{1,i_1}\ and\ \ldots\ and\ x_n\ is\ A_{n,i_n}\ THEN\ y\ is\ B_{i_1,\ldots,i_n},$$

where A_{k,i_k} is the antecedent i_k belongs to input k, $B_{i_1,\ldots,i_n}$ is the fuzzy set that is assigned to the consequent set of the rules, $i_j = 1, \ldots, n_j$; and n_j is the number of the antecedent sets belonging to input j. The rule premises can be obtained from all the possible combinations of the fuzzified values of the inputs.

2.1 Fuzzification

The system inputs are generally obtained as crisp values. In order to handle inherent uncertainty, a fuzzy value should be made from them by applying the fuzzification method. In this step the degree of fit of the input should be

determined on the antecedent set, i.e. the extent to which current inputs belong to the fuzzy sets, which are used to characterize them [8]. In the case of triangular membership functions and crisp inputs, the fuzzified value can be calculated using Eq. (1) and it is illustrated in Fig. 1.

$$\mu_{A_i}(x)=\begin{cases} 0 & if \quad x \le a \\ \dfrac{x-a}{b-a} & if \quad a \le x \le b \\ \dfrac{c-x}{c-b} & if \quad b \le x \le c \\ 0 & if \quad c \le x \end{cases} \tag{1}$$

where a_i, b_i, c_i are the membership function parameters. Visually it looks like this:

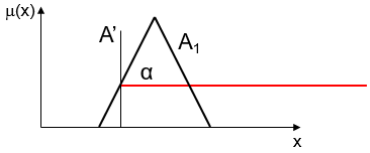

Figure 1
Fuzzification for triangular membership function

2.2 Firing Strength Calculation

During the evaluation the fuzzified values of all the input factors should be taken into account. For this reason, these values should be connected by a fuzzy operator for each rule, based on its antecedent to get the rule strength, which specifies the influence level of the given rule antecedent [9]. The operator selection depends on the task. In the case of a product operator, it can be calculated using Eq. (2) and its result is shown in Fig. 2. The product formula is

$$w_i = \prod_{j=1}^{m} \mu_{A_{kj}}(x) \tag{2}$$

where m is the number of the inputs, $\mu_{A_i}(x)$ is the fuzzified value of antecedent k of input j. And a visualization of it is

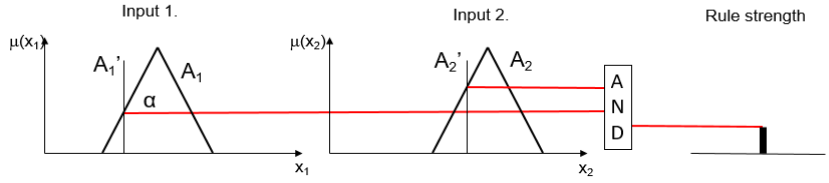

Figure 2
Firing strength calculation using product operator

2.3 Fuzzy Implication

After the rule strength has been determined for each rule we need to find the extent to which the rule-consequent is involved in the final result [10]. Different operators can be used depending on the task. In Fig. 3. the result of a product implication can be seen, which can be calculated using

$$y_{B_i} = w_i \mu_{B_i} \tag{3}$$

where w_i is the firing strength of rule i and μ_{Y_i} is the consequent set belonging to rule i. Visually, it looks like this:

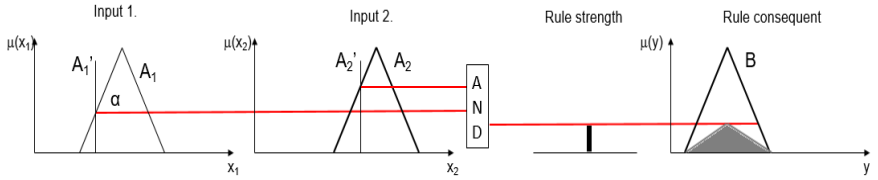

Figure 3
Fuzzy implication using product operator

2.4 Aggregation

After the implication has been calculated for each rule, the partial conclusions are obtained. These fuzzy sets should be aggregated via Eq. (4) to get the system result using an averaging operator [11]. Fig. 4 shows an example using the aggregation method with a sum operator. If the membership functions form a Ruspini-partition, i.e. $\sum_{i=1}^{n} \mu_i(x) = 1$, then it is not necessary to weight the sets obtained by using the firing strength [12].

Here,

$$y = \sum_{i=1}^{n} w_i y_{B_i} \tag{4}$$

where w_i is the firing strength of the rule-premise i, y_{B_i} is the result of the implication method for rule i, and n is the number of rules. Visually, it looks like

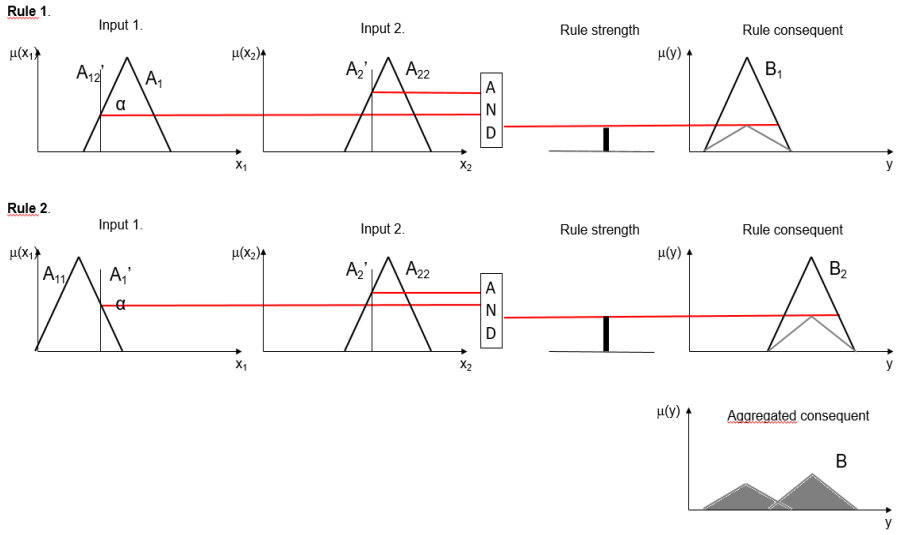

Figure 4
Sum aggregation

2.5 Defuzzification

In the case where a crisp output is needed, which characterizes the system result best, the complex shaped function obtained should be defuzzified. There are several different defuzzification methods available for, of which the centroid method is the most common and physically appealing of all the defuzzification methods [13]. Fig. 5 shows the results of the defuzzification process using the centre of gravity method, which can be defined by

$$y_{out} = \frac{\int_{y \in \mu_B} \mu_B(y)\, y dy}{\int_{y \in \mu_B} \mu_B(y) dy} \tag{5}$$

where $\mu_B(y)$ is the aggregated consequent set.

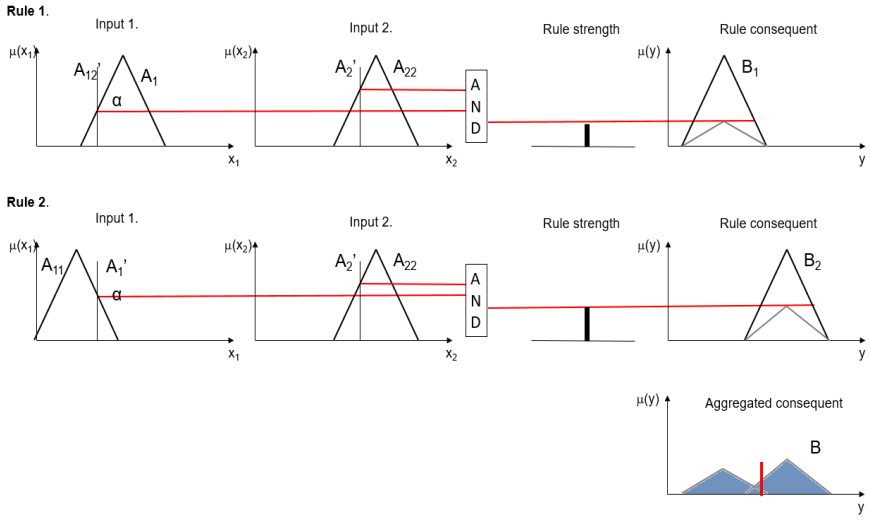

Figure 5
Centre of gravity (COG) defuzzification

3 Mamdani-like System with Discretized Output

3.1 Basic Principles of the Modifications

In the original Mamdani model aggregation and defuzzification are the evaluation steps, which increase the computational needs to the greatest extent. The idea of our reduction approach came from this fact. The first part of the evaluation (i.e. fuzzification, firing strength calculation, fuzzy implication) is the same as that in the original model, but the order of the aggregation and defuzzification is swapped. After the fuzzy implication step, each rule-consequent set is first defuzzified, then it is followed by aggregation. This modification might appear to further increase computational needs, because defuzzification is the computationally most intensive task and in this version, it should be performed as many times for as many rules there are. However, the high computational needs of the defuzzification in the original model is caused by the complex shape of the function, which is obtained after the aggregation. In contrast, in the modified model, simple shaped functions should be defuzzified, whose computational needs are negligible. And, during the aggregation significantly fewer operations are required because in this case crisp values should be aggregated instead of fuzzy sets. Detailed properties and a proof of the equivalency can be found in [17].

3.2 Defuzzification

In this modified version of the original Mamdani-type control, defuzzification is performed for each rule consequent set separately. These sets are piecewise linear (i.e. triangular, trapezoidal) fuzzy sets. As can be seen from Eq. (6) this simplified form significantly reduces the computational needs, for a triangular membership function.

Here,

$$y_i = \frac{a_i + b_i + c_i}{3} \tag{6}$$

where a_i, b_i, c_i, are the membership function parameters.

In the case of a symmetric function, the calculation is even simple, because in the case of triangular function the defuzzified value is just equal to b_i. In the case of symmetric trapezoidal function, it can be calculated using

$$y_i = \frac{b_i + c_i}{2} \tag{7}$$

The defuzzification method for the simplified model is shown in Fig. 6 below.

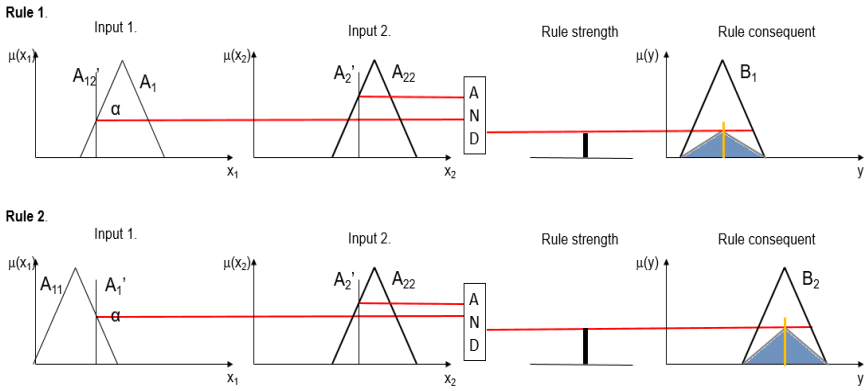

Figure 6
Defuzzification in the Mamdani-like system with discretized output

3.3 Aggregation

After defuzzified values have been obtained for all the rules, the final conclusion of the system should be drawn. For this reason, these crisp values should be aggregated as a weighted sum of them, where the weight factor is the firing strength of the rules, as follows

$$y_{out} = \frac{\sum_{i=1}^{n} w_i y_i}{\sum_{i=1}^{n} w_i} \tag{8}$$

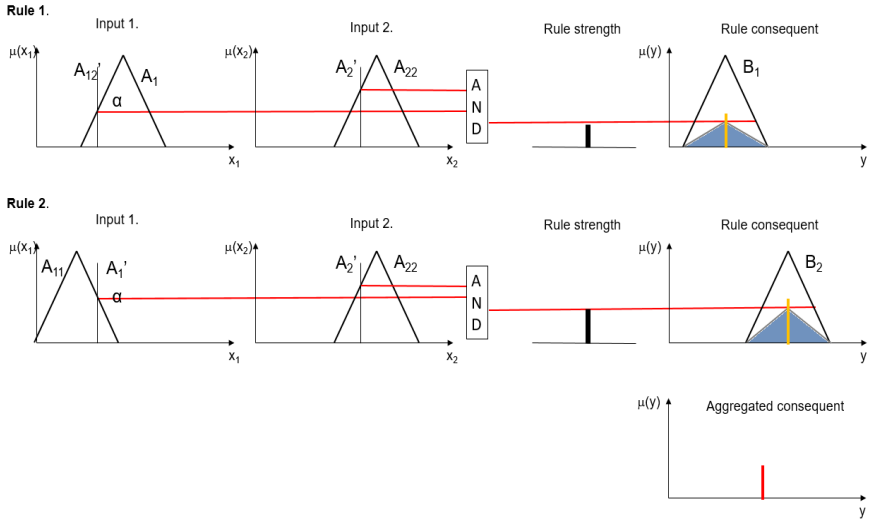

Figure 7
Aggregation in the Mamdani-like system with discretized output

4 Arithmetic-based Control

4.1 Basic Principles of the Method

The traditional Mamdani-type control is again used here as a basis for our simplification approach. Even the order of the evaluation steps is the same as in the traditional model (Subsection 2.1-2.5); only the step of fuzzy implication is omitted, while aggregation is modified. The basic idea is to manage the left and right-hand sides of the functions instead of aggregating the entire consequent sets. In this case, the result of the aggregation can be got as the linear combination of these parts of the functions, left and right-hand sides separately [14]. Consequently, the computational needs can be significantly reduced, due to using crisp parameters to represent the membership functions. The proposed method is equivalent to the α-cut, but it is more effective. Detailed properties and a proof can be found in [15].

4.1 Aggregation

After the firing strength calculation (2), the values obtained should be normalized like so:

$$w_{i_n} = \frac{w_i}{\sum_{i=1}^{n} w_i} \tag{9}$$

The left- end right-hand side of the aggregated consequent set can be represented by the following pair of equations

$$y_L = m_L(x - a_L) + \frac{1}{2} \tag{10}$$

$$y_R = m_R(x - a_R) + \frac{1}{2} \tag{11}$$

where m_L and m_R are the slopes of the membership function, $m_L > 0, m_R < 0$, $a \in X$, and $f(a) = \frac{1}{2}$.

A linear combination of the rule consequences can be obtained, for both sides of the functions, as follows:

$$a = \sum_{i=1}^{n} w_{i_n} a_i \tag{12}$$

$$\frac{1}{m} = \sum_{i=1}^{n} \frac{w_{i_n}}{m_i} \tag{13}$$

After the aggregation, the simple-shaped function obtained can be defuzzified using Eq. (6). The way of aggregation and defuzzification work is shown in Fig. 8 below.

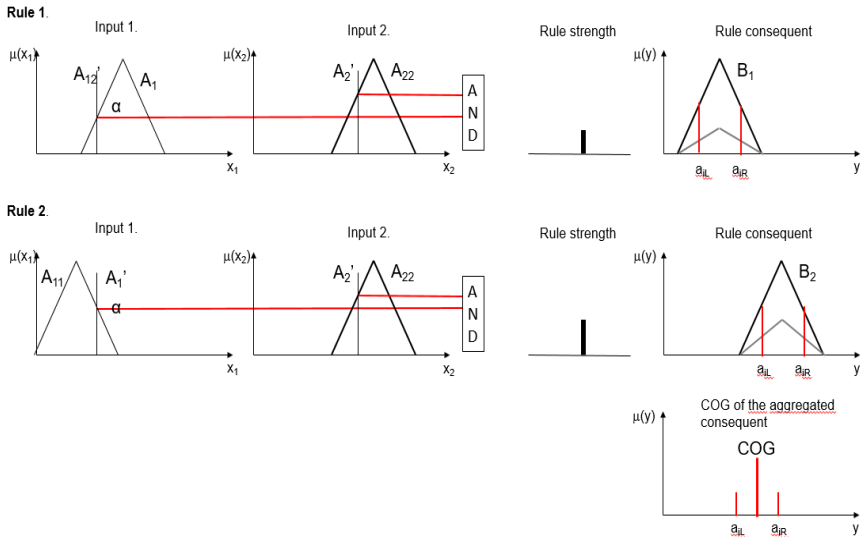

Figure 8

Aggregation and defuzzification in the arithmetic-based control process

5 Comparison of the Computational Needs

5.1 Mamdani-type Control vs. Mamdani-like System with a Discretized Output

When comparing the computational needs of these methods, it should be kept in mind that the first steps of the evaluation (i.e. fuzzification, firing strength calculation, implication) are the same. Consequently, their computational needs do not have to be taken into account because they are the same. This significant reduction can be achieved due to the swapping of the order of the aggregation and defuzzification.

In the traditional Mamdani-type model, fuzzy sets should be aggregated, then the complex shaped function obtained should be defuzzified. In practice, this means that aggregation must be performed with an equidistant division of the input range. Let Y be the input domain $[y_{min}, y_{max}]$, the set, which contains its equidistant base points $Y_i = [y_1; y_1 + \Delta; y_1 + 2\Delta; \dots; y_N]$, where N is the number of the base points of the equidistant division and the distance between two points is $\Delta = (y_{max} - y_{min})/(N-1)$. Aggregation should be performed for

these points using Eq. (4), thus the number of operations requirements of this step is $N * n$ multiplicative and $n - 1$ additive operations. During the defuzzification, the trapezoidal rule is applied like so:

$$\int_{y_{\min}}^{y_{\max}} f(y)dy \approx T_n(f) = \frac{y_{\max} - y_{\min}}{2N}\left[f(y_1) + 2\sum_{i=2}^{N} f(y_i) + f(y_{N+1})\right] \tag{24}$$

where $y_i = y_1 + (i-1)\frac{y_{max} - y_{min}}{N}$, *i=1, …, n+1*. The number of operations needed for the COG method, which is defined in (5) is *N+1* additive and 3 multiplicative operations (since aggregated values are given in the base points). This is equal to the computational needs of the denominator of the formula, while in the nominator *N+1* additive and *N+3* multiplicative operations are needed. Hence, the overall formula requires *2(N+1)* additive and *N+6* multiplicative operations.

In the Mamdani-like structure with discretized output, the centre of gravity of each simple-shaped rule consequent is calculated. Using Eq. (6) the number of operations needed is *2* additive, and *1* multiplicative operation for each rule consequent. Then, these crisp values should be aggregated as their weighted sum, where the weight factor is the firing strength of the rules defined by Eq. (8). Its computational needs are $2(n-1)$ additive, and *n* multiplicative operations.

A summary of the above is shown in Table 1.

Table 1
Comparison of the computational needs for the traditional and discretized Mamdani-type control processes

Operation		Traditional Mamdani	Discretized Mamdani
Aggregation	Additive	n-1	2n
	Multiplicative	N*n	n
Defuzzification	Additive	2(N+1)	2(n-1)
	Multiplicative	N+6	n
Sum	Additive	n-1+2(N+1)	4n-2
	Multiplicative	N*(n+1)+6	2n

Although the Takagi-Sugeno system was not addressed, a brief comparison was made for the sake of completeness regarding the computational requirements. The results are shown in Table 2. Takagi-Sugeno is a standard method, where the rule consequents can be generated as a function of input values using $g_{i_1,\ldots,i_n}(x_1,\ldots,x_n)$, and the rules can be represented in the following form [18]:

$$x_1 \text{ is } A_{1i_1} \text{ and } \ldots \text{ and } x_n \text{ is } A_{ni_n} \text{ THEN } y \text{ is } g_{i_1,\ldots,i_n}(x_1,\ldots,x_n).$$

As a result of the above the computational requirements of the function should be examined, instead of the defuzzification step, which is omitted.

Table 2
Comparison of the computational needs for the Takagi-Sugeno and discretized Mamdani-type control processes

Operation	Takagi Sugeno	Operation type	Discretized Mamdani	Operation
Aggregation	2(n-1)	Additive	2n	Aggregation
	n	Multiplicative	n	
Output function	2n	Additive	2(n-1)	Defuzzification
	3n	Multiplicative	n	
Sum	4n-2	Additive	4n-2	Sum
	4n	Multiplicative	2n	

The results in the table clearly show that the number of operations needed for the traditional Mamdani control system is significantly higher than in the case of its discretized version, for both additive and multiplicative operations.

5.2 Mamdani-like Structure with Discretized Output vs. Arithmetic-based Control

Comparing the two methods, it should be mentioned that in the arithmetic-based algorithm fuzzy implication is not a separate step, and the firing strength is used during the aggregation process to weight the parameters of the functions. For this reason, the number of operations needed for fuzzy implication cannot be defined for this approach. Another important issue is defuzzification. In both cases, simple-shaped functions should be defuzzified. As a consequence, the number of operations needed for defuzzification do not have to be taken into account because their computational needs are the same.

In the case a Mamdani-like system with a discretized output, during the implication the above-defined equidistant division is used. The function values should be multiplied by the firing strength of the given rule (3), hence it requires N*n multiplicative operations. In the arithmetic-based method, we also need to aggregate crisp values as in the discretized Mamdani model. However, here $a \in X,\ f(a) = \frac{1}{2}$ (12), and the slope of the functions (13) should be aggregated. Based on these equations, the computational needs of the aggregation are $4(n-1)$ additive, and *4n* multiplicative operations.

After summarizing the operational needs of the methods, it can be seen that the arithmetic-based method provides a slightly worse result for additive operations,

but discretized Mamdani structure has significantly higher computational needs for multiplicative operations. It may be deduced from the above that the arithmetic-based method has fewer computational needs overall.

Table 3
Comparison of the computational needs of the discretized Mamdani-type and arithmetic-based control

Operation		Discretized Mamdani	Arithmetic-based
Implication	Additive	-	-
	Multiplicative	N*n	-
Aggregation	Additive	2n	4(n-1)
	Multiplicative	n	4n
Sum	Additive	2n	4(n-1)
	Multiplicative	(N+1)n	4n

6 A Case Study

The above-described methods were examined in a patient surveillance system, where physiological parameters were monitored in real-time. In these kinds of systems, a short reaction time is essential to avoid any serious consequences caused by the inappropriately chosen form, or intensity of the motion [16]. In real-time systems use of conventional Mamdani control is limited, due to its high computational requirements. However, with a modified algorithm one can retain its advantages, and also ensure a faster reaction time.

6.1 Model Structure

In the system, input parameters are selected patient-by-patient, based on a personal profile. This means that the number of the monitored parameters and their ranges can be defined specifically for patients and sports. A schematic structure of the model is shown in Fig. 9. The aim of the system is to assess the risk level of the current activity, based on the measured physiological values in the "Medical condition" group; the characteristics of the sport activity are its duration, frequency, and intensity under the "Activity" group; and in the case of an outdoor activity, the environmental conditions are listed in the third main group. An evaluation is performed hierarchically, i.e. risk levels of the different groups are calculated separately for each group, then these values form the input parameters of the summarizing group at the next level, where the overall risk is obtained. Due to the hierarchical structure, the system can be easily modified or expanded, where necessary.

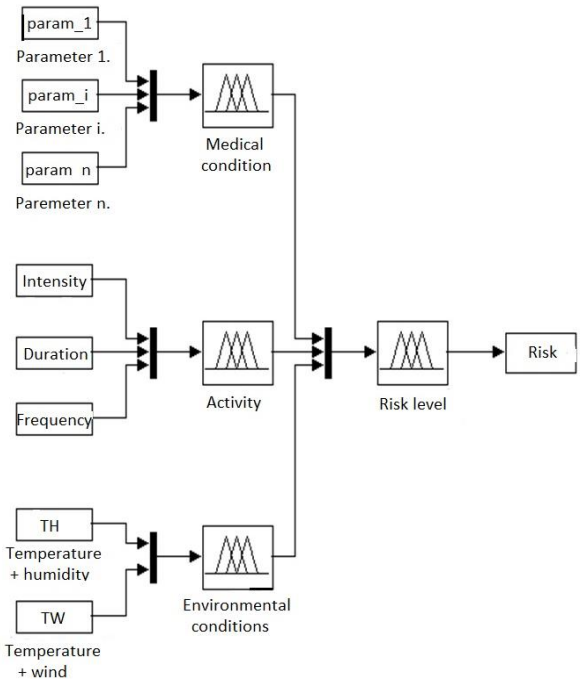

Figure 9
Model structure

6.2 A Comparison of the Control Models

During the study the heart rate, systolic and diastolic blood pressure of the patients were examined for 576 different input value combinations, for a 50-year old female patient.

First, the results of the different algorithms were compared to demonstarte the interchangeability of each. The compatibility of the conventional Mamdani control and the Mamdani-like control with a discretized output has already been shown mathematically [17]. For the two discretized solutions, the results of Mamdani-like structure and arithmetic-base model are shown in Fig. 10. Quantifying the comparison of the result: the average difference, the maximum difference, the standard deviation, and the correlation-coefficient were also calculated (see in Table 4). The figure and the metrics tell us that the output values in both systems are almost the same, the difference between them being negligible, supporting the former mathematical proof.

Here, the computational speed of the different methods is also considered in a standard desktop environment. The algorithms were implemented in a RAD

Studio multi-platform, multi-device environment. The running time for the 576 different cases are as follows: 8.017 s for the conventional Mamdani control, and 4.309 s for Mamdani-like control, 4.283 s for the arithmetic-based model. This means that the average evaluation time of the current input for the different algorithms is 14 ms, 7.5 ms, and 7.4 ms, respectively. From the above, we may infer that the computational speed is approximately the same as in the case of the Mamdani-like control and arithmetic-based models, while the conventional Mamdani control requires almost twice as much CPU time. As can also be seen from the mathematically described computational requirement (Table 1, 3).

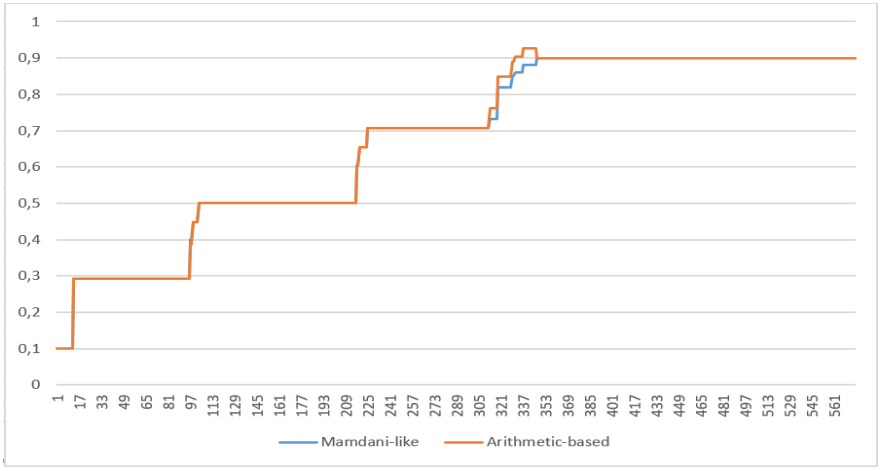

Figure 10
Comparison of the Mamdani-like and arithmetic-based control results

Table 4
Comparison of the results of the discretized Mamdani-type and arithmetic-based control for 576 different cases

Average difference	0.0022
Maximum difference	0.0455
Standard deviation	0.0091
Correlation-coefficient	0.9993

Conclusions

Fuzzy control systems are widely used in medical-related applications because of their advantageous properties. That is, linguistic terms can be incorporated into the model, it can handle the subjectivity in the data and in the evaluation process, and its operation is very close to human thinking. However, the applicability of the conventional Mamdani control is limited by its high computational requirements in real-time and adaptive systems. In this study, possible modifications of the conventional Mamdani model were suggested, such as the

Mamdani-like structure with a discretized output and the arithmetic-based model. The main contribution of the paper is that the proposed models maintain the basic "philosophy" of the Mamdani-type controller while utilizing the properties of the triangular and trapezoidal membership functions technical modifications were suggested that drastically reduce the computational needs in comparison with that of the original method that is formulated for general form membership functions. Computational requirements of the different approaches were compared mathematically in general, for any-shape function. Besides that computational speed was also compared in a patient monitoring system. Based on the analysis both the Mamdani-like system and arithmetic-based model have lower computational needs than the conventional Mamdani model. There is no significant difference between the two discretized models, and both can enhance the features of the conventional model by decreasing the computational requirements and CPU time required.

References

[1] T. Obeidi, C. Larbes, A. Ilinca,G. F.T. Kebir, Fuzzy Logic-based Maximum Power Point Tracking fora Solar Electric Vehicle, Acta Polytechnica Hungarica, Vol. 15, No. 7, pp. 133-156, 2018

[2] A. Gegov, "Complexity management in fuzzy systems," *Chapter Rule Base Reduction Methods of Fuzzy Systems of Series in Studies in Fuzzyness and Soft Computing*, Springer, Heidelberg, pp. 17-31, 2007

[3] R-E. Precup, "On the combination of tensor product and fuzzy models," In Proc. of the IEEE Int. Conference on Automation, Quality and Testing, Robotics, Cluj-Napoca, 22-25 May, pp. 48-53, 2008, DOI: 10.1109/AQTR.2008.4588792

[4] F. Herrera, "Genetic fuzzy systems: taxonomy, current research trends and prospects," Evolutionary Intelligence, Vol. 1, No. 1, pp. 27-46, Springer-Verlag, 2008

[5] S. Chopra, R. Mitra, V. Kumar, "Reduction of fuzzy rules and membership functions and its application to fuzzy PI and PD type controllers," Int. Journal of Control, Automation, and Systems, Vol. 4, No. 4, pp. 438-447, 2006

[6] S. Ding, H. Zhao, et. al, "Extreme Learning Machine: Algorithm, Theory and Applications", Artificial Intelligence Review, Vol. 44, No. 1, pp. 103-115, June 2015

[7] Y. Chen, A. Abraham, "Hierarchical fuzzy systems," Tree-Structure based Hybrid Computational Intelligence, Springer, Heidelberg, Vol. 2, 2010, pp. 129-147

[8] T. J. Ross, Fuzzy Logic with Engineering Applications, Second Edition, John Wiley &Sons, 2004

[9] R. Czabanski, M. Jezewski, J. Leski, Introduction to Fuzzy Systems, Theory and Applications of Ordered Fuzzy Numbers, Studies in Fuzziness and Soft Computing, Vol. 356, pp. 23-42, 2017

[10] Kyung WhanOh, WyllisBandle, Properties of Fuzzy Implication Operators, International Journal of Approximate Reasoning, Vol. 1, No. 3, pp. 273-285, 1987

[11] R. Fullér, Aggregation Operators – Tutorial, Institute for Advanced Management Systems Research Department of Information Technologies Abo Akademi University, 2010

[12] A. R, Várkonyi-Kóczy, "Model Based Anytime Soft Computing Approaches in Engineering Applications." In Balas, V., J. Fodor, A. R. Várkonyi-Kóczy (eds.), Soft Computing Based Modeling in Intelligent Systems (Ser. Studies in Computational Intelligence), Springer Verlag, Berlin, Heidelberg, 2009, pp. 63-92, DOI: 10.1007/978-3-642-00448-3_4

[13] Oļegs Užga-Rebrovs, Gaļina Kuļešova, Comparative Analysis of Fuzzy Set Defuzzification Methods in the Context of Ecological Risk Assessment, Information Technology and Management Science, Vol. 20, No. 1, pp. 25-29, 2018

[14] J. Dombi, T. Szépe, Arithmetic-based Fuzzy Control, Iranian Journal of Fuzzy Systems, Vol. 14, No. 4, pp. 51-66, 2017

[15] J. Dombi, Pliant arithmetics and pliant arithmetic operations, Acta Polytechnica Hungarica, Vol. 6, No. 5, 2009, 19-49

[16] Ábel Garai, István Péntek, Attila Adamkó, Revolutionizing Healthcare with IoT and Cognitive, Cloud-based Telemedicine, Acta Polytechnica Hungarica, Vol. 16, No. 2, pp. 163-181, 2019

[17] E. Tóth-Laufer, I. J. Rudas, M. Takács, Operator Dependent Variations of the Mamdani-type Inference System Model to Reduce the Computational Needs in Real-Time Evaluation, International Journal of Fuzzy Systems, Vol. 16, No. 1, pp. 57-72, 2014

[18] T. Takagi, M. Sugeo, Fuzzy identifcation of systems and its applications to modeling and control, IEEE Transactions on System, Man and Cybernetics, Vol. 15, pp. 116-132, 1985

Ranking and Assessment of the Efficiency of Social Media using the Fuzzy AHP-Z Number Model - Fuzzy MABAC

Zoran Bobar[1], Darko Božanić[2], Ksenija Djurić[2], Dragan Pamučar[2]

[1] Ministry of Defense of the Republic of Serbia, Bircaninova 5, 11000 Belgrade,
[2] Military Academy, University of Defence in Belgrade, Generala Pavla Jurisica Sturma 33, 11000 Belgrade, Serbia
emails: zoran.bobar@mod.gov.rs, darko.bozanic@va.mod.gov.rs, ksenija.djuric@mod.gov.rs, dragan.pamucar@va.mod.gov.rs

Abstract: Methods of ranking and evaluation of the effectiveness of Social Media (SM) given and applied in this paper are the basis for the selection of online media that the public administration uses when communicating with citizens. The methodology presented is based on multicircular decision-making using the Fuzzy Analytical Hierarchical Process (fuzzy AHP) - Z number model - Fuzzy Multi-Attributive Border Approximation Area Comparison (fuzzy MABAC), which eliminates the traditional intuitive ratings of PR services. This resulted in poor use of available channels of communication or ineffective communication. Positive results of the application of presented methods are especially evident in increasing the number of channels of communication on the Internet and the realization of communication goals for greater participation of citizens in public administration.

Keywords: Internet; Social Media; Fuzzy Analytical Hierarchical Process (fuzzy AHP); Fuzzy Multi-Attributive Border Approximation Area Comparison (fuzzy MABAC); Z-number

1 Introduction

Governmental institutions, businesses and other legal entities are increasingly using social media as channels of communication with their target public. Until recently it was common practice to have PR tools located on the front page of the site in the form of news, multimedia, contacts, and links, serving as a convenient way of servicing classic media [1]. More recent research has revealed that the public is spending an increasing amount of time on social media, and that the potential of social media is based on interactions, collaboration and cooperation,

where social media is seen to transform the relationship of public service and the citizenry [2]. In the process of communication with public, government institutions and businesses are not able to use the virtually unlimited number of media available on the Internet with equal efficiency. The selection of social media to be used as a communication channel ensuring reach amongst the targeted public online with desired efficiency is one of the challenges faced by these services in implementing a given communication strategy.

In line with the new trends in the usage of social media for business purposes, a series of studies have examined the justification of using social media in the communication of public administration with the citizenry. Some focused on functional availability [3], benefits and risks [4], and the impact of social media use [5]. The common trait found in all these studies is the analysis and assessment of multiple social media according to a single criterion, or the analysis and evaluation of a single social media based on multiple criteria. However, communicators are often confronted with the need to choose among available social media. They have to identify the criteria and the methodology for making such a choice to reduce the risk of failure in communicating via social media.

Considering the trends in social media usage and the perspective of involving such media in the activities of public administration as well as that of other organizations, the aim of this research is:

1. To define the general criteria for choosing social media for communication between government bodies and the target public; and
2. To develop the methodology for social media ranking in view of the selection of a particular communication channel suited to a communication strategy.

This paper examines the various options for rating and ranking social media by using parameters that influence the efficiency of communication, defined on the grounds of criteria partially used in the evaluation and ranking of social media, as well as on expert assessments of the applicability of such criteria.

2 Review of Literature

Although practically all state institutions have websites today, the needs of citizens using the Internet have not entirely been met. The emergence of Web 2.0 technology prompted the development of new public administration services for the implementation of eGov 2.0 or Social Government projects [6] [7] which enable an interactive dimension of public administration through online activities.

The approach most frequently used in the development of eGovernment is the evolutionary one. Namely, public administration portals evolve from a basic

presence on the Internet, through interaction and then to collaboration with citizens [8]. Public relations through the web portal of the institution gradually evolve from one-way communication, i.e. from providing information on websites to the use of online services for two-way communication and connections with institutional sites and profiles on social media (Social Network Room). This implies a high-level of interaction with the target audience [9]. Social media enable greater participation of citizens and stakeholders, create a space for cooperation with government and stimulate innovation in the public sector [10]. For the public sector, communication through social media is more targeted, more economical and more efficient compared to other communication channels. If the information is available on social media, reactions can be expected quickly. Such feedback can help speed up more efficient decision-making and serve as a way of confirming public support. Feedback also reveals which media should be used as carriers of specific information in communicating with the public [11]. Citizens and public authorities have different expectations regarding the use of social media. Citizens expect their government to use social media as a tool which would give them a stronger voice in the decision-making process. Conversely, the administration sees social media as a tool for interacting with citizens, and a way of promoting services in line with current government policies and strategies. [12].

In using social media to communicate with citizens, the public sector pays insufficient attention to the "variants among social media tools, i.e., which tools are appropriate for certain activities and which for others". Although the adoption of new technologies is important, it is equally important to assess how these technologies are used and to which purpose. Some technologies are broadly used, but not with their full potential. The problem faced by the public sector is the excessive use of social media by employees, as this can generate problems in managing communications or ensuring that all communications are true and consistent [13].

2.1 Review of Criteria for Evaluation of Social Media Efficiency

The initial assumption of this paper was the fact that the use of social media is determined by a number of factors. Thus, the study recruited 35 researchers working in the field of public administration and public relations. Based on the evaluation of their competence, 11 experts were selected. Experts were polled in more than one round using the Delphi method. Criteria presented in the literature were offered to the experts, along with the possibility of adding new ones and eliminating the ones already offered. In this research further criteria for the assessment of SM efficiency in communication with society were proposed:

Criterion 1 (C1): Functionality. Social media can be evaluated based on available tools classified into functional blocks: identity, conversation, sharing, presence, relationships, reputation and groups [14]. The classification of social media in

relation to the availability of functional blocks enables us to look at different levels of presence on the Internet and networking, or how and to what level of availability can a given social medium be configured. The convenience of functional classification lies in the ability to compare existing social media and assess their capabilities in relation to the target audience.

Criterion 2 (C2): Presence. Social media can be present in a certain geographical area to a greater or lesser extent. If the targeted public is in that area, the influence of such presence on the selection of whether and to what extent these particular social media platforms be used is related to the share of the targeted public using that medium. Presence, measured by the number of users of the social network, also affects the level of reach of the targeted public, although it does not have to be directly proportional to the number of users of that platform. Quantitative data for this criterion are reliable and are found as statistical data for each social medium [15].

Criterion 3 (C3): Purpose. Social media can also be evaluated according to their purpose because the functional richness of a social medium is not sufficient to assess the effectiveness of that medium in relation to individual social groups. In general, once social media find their own community their true purpose is revealed. The information offered and the way of communicating are most often suited to a particular type of media. Kaplan et al. categorized social media by their purpose as follows: Collaborative Projects, Blogs and Microblogs, Content Communities, Social Networking Sites, Virtual Game Worlds, Virtual Social Worlds [16].

Criterion 4 (C4): Richness of content. The effectiveness of social media can also be viewed through the richness of media content described in Media Content Theory [17]. Modern public relations theory is based on two-way communication with the target audience, which implies the use of social media precisely because of its interactive nature. Social media research suggests the need to expand Media Content Theory, taking into account the phenomenon of social networking and communication efficiency achieved through massive networking and the collaboration of communicators through those media [18].

Criterion 5 (C5): Target audience. It is customary for the same target group to use several social media platforms. One of the ways in which the target audience is determined is to rely on available statistics of a social media platform or on relevant databases offering information on the use of social media. Media coverage and target audience do not need to be directly linked, and significant deviations in social media are possible. For example, although Facebook is used by more than 47.6% of the population of the Republic of Serbia, the percentage of those aged over 60 is only 10%, and male population percentage is higher [19]. Therefore, the target audience on a social media platform must be segmented, in order to assess its suitability for communication with the public.

Criterion 6 (C6): Engagement., Despite the functional availability of various tools of social media, the public recognizes some as more or less effective or credible regarding the content provided. Surveys covering social network users indicate multiple levels of engagement (liking, sharing, comments), which is directly related to the level of trust in that media [20].

Based on available literature and the positions agreed on by the expert group, an overview of the criteria that influence the choice of a social media platform for public administration communication with the citizens has been formed, as indicated in Table 1.

Table 1
Criterion – basic characteristics

Criterion	Description	Type	Value trend	Range
C1	Functionality	Quantitative	Growing	1 to 7
C2	Presence	Quantitative	Growing	1 to 5
C3	Purpose	Qualitative	Growing	1 to 5
C4	Content richness	Qualitative	Growing	1 to 5
C5	Target audience	Quantitative	Growing	1 to 100%
C6	Engagement	Qualitative	Growing	1 to 5

3 Method Description

A specific research method has been developed to evaluate and rank the effectiveness of social media. The model is defined based on examination of the literature related to selection of social media platforms and mulitple-criteria decision making implemented in other research fields. The authenticity of the criteria based on which the ranking is made, as well as the inter-relations of the criteria, conditioned the implementation of AHP method, as a method frequently used in research. During the construction of the model, the experts showed certainin accuracy when comparing pairs which is the requirement of the AHP method. With that regard the fuzzyfication of the method was performed. Similar output occurred with the MABAC method, which is emphasized as one of the methods which yields highly significant results. However, the course of examination showed that the experts are often hesitant when assigning the values to those criteria where there is no exact data. Hence, the degree of confidence was introduced in the ranking process, which was later mathematically processed.

When defining the method which should be used, the main criteria were the method stability and the ease of confirming the results. The stages and steps of the model are defined in Figure 1. When defining the method which should be used, the main criteria were the method stability and the ease of confirming the results.

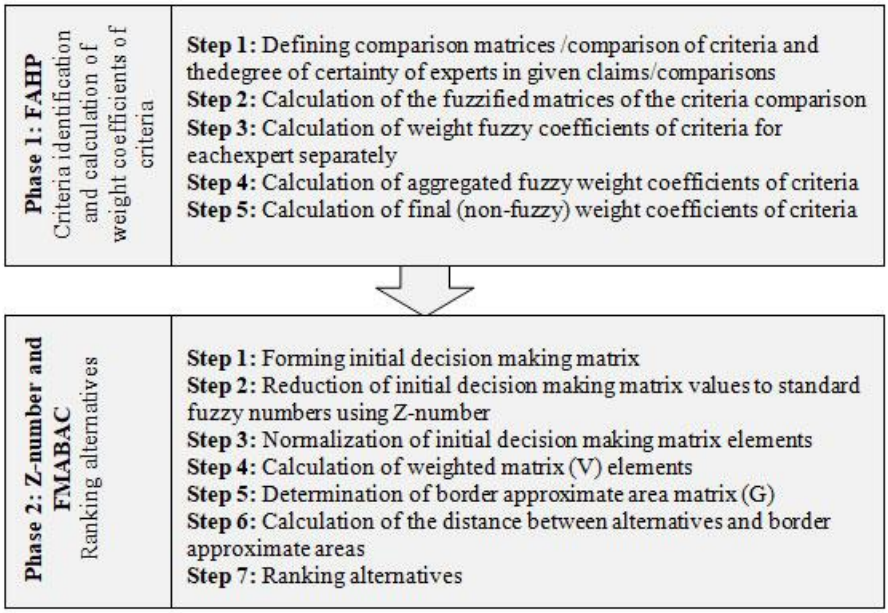

Figure 1
The stages and steps of the model

The used methods are described in the next part of the paper.

3.1 Fuzzy AHP

An Analytical Hierarchical Process (AHP) is a method often used in multi-criteria decision-making. This method was developed by Thomas Saaty. It is based on the explanation of a complex problem in the hierarchy, with the aim at the top, criteria, sub-criteria, and alternatives at the levels and subjections of the hierarchy [21]. Apart from the hierarchical structuring of the problem, AHP differs methodologically from other multiple-criteria methods, due to its comparisons of elements in pair on the given level of hierarchy with the elements on the higher level of hierarchy. For each of the elements at the higher level $n(n-1)/2$ a comparison of semantic or numerical type, as defined with Saaty scale is needed,

Table 2. Saaty's scale is used to determine the weight coefficients of the criteria, as well as to rank the alternatives.

Table 2
Saaty's scale for comparison in pairs

Standard Values	Definition	Inverse values
1	The same importance	1
3	Week dominance	1/3
5	Strong dominance	1/5
7	Very strong dominance	1/7
9	Absolute dominance	1/9
2, 4, 6, 8	Intervalues	1/2, 1/4, 1/6, 1/8

The comparison of elements on the given level of hierarchy yields the matrix of comparisons A:

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ . & . & \dots & . \\ . & . & \dots & . \\ a_{n1} & a_{n1} & \dots & a_{nn} \end{bmatrix} \quad (1)$$

The calculation of the weight coefficients, or in other words the ranking of alternatives, is related to different prioritizing approaches. Respectively, the most common approaches are the Eigenvector method (EV), Logarithmic Least Squares method (LLS), and Additive Normalization method (AN). In that regard, this research used the Additive Normalization method (AN).

The calculation of the priority vector w is gained through 3 steps:

Step 1: each element from the given matrix column A is divided with the sum of elements of that column;

Step 2: all elements in each column are gathered

Step 3: each of the given sums is divided with the range of matrix n

This procedure is described with visuals (2) and (3):

$$a_{ij}^{'} = \frac{a_{ij}}{\sum_{i=1}^{n} a_{ij}}, \quad i, j = 1, 2, ..., n \quad (2)$$

$$w_i = \frac{\sum_{i=1}^{n} w_{ij}}{n}, \quad i = 1, 2, ..., n \quad (3)$$

Very often Saaty's scale is being modified with the application of fuzzy sets.

The basic items of the fuzzy logic were given by Lotfi Zadeh [22], [23], [24]. The basics that L Zadeh gave were sufficient to continue to develop fuzzy logic and to be increasingly applied in practice. In this paper, we use the triangular fuzzy numbers $T = (t_1, t_2, t_3)$, Figure 2, where t_1 represents the left, and t_3 the right distribution of the confidence interval of the fuzzy number T and t_2, where the function of fuzzy number membership has a maximum value, one.

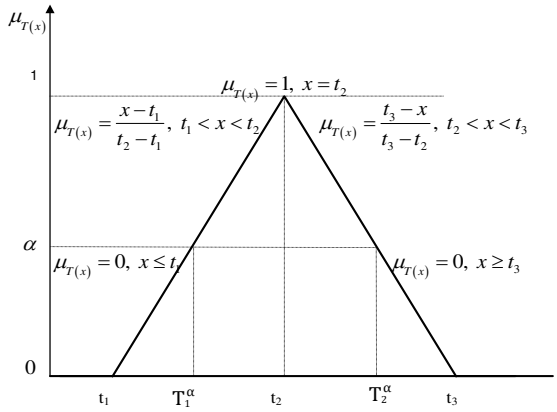

Figure 2

Triangular fuzzy number T [25]

At the end of the application, the fuzzy number is converted to a real number. A number of methods are used for this procedure [26]. Some of the known defazification expressions are [27] [28]:

$$A = ((t_3 - t_1) + (t_2 - t_1))/3 + t_1 \tag{4}$$

$$A = \left[\lambda t_3 + t_2 + (1-\lambda) t_1\right]/2 \tag{5}$$

where λ represents an index of optimism, which can be described as a belief/decision maker's relationship to decision-making risk [29]. The most common optimism index is 0, 0.5 or 1, which corresponds to the pessimistic, average or optimistic view of the decision maker [29].

There are different approaches to fuzzification of Saaty's scale. In principle, they can be divided into two groups: sharp (hard) and soft fuzzification [30]. Under the "sharp" fuzzification, it is assumed that for a fuzzy number $T= (t_1,t_2,t_3)$ confidence interval is predetermined, i.e., it is defined in advance that the value of the fuzzy number will not be greater than t_3 nor less than t_1 [30]. In soft fuzzification, the confidence interval is not predetermined, but it is defined during the decision-making process, based on additional parameters.

A common occurrence is the definition of a scale with a smaller number of comparisons than Saaty's Six-step scale [31] [32] [33], while a five-step scale is used in Mandic et al, [34], Junior et al. [35] and Ruiz-Padillo et al. [36]. Most of

these scales use the model of the fuzzy number $T = (t_1,t_2,t_3) = (x-1,x,x+1)$, where x represents the standard value of Saaty's scale. Some authors modifies the Saaty scale use the model of the fuzzy number $T=(x-\delta,x,x+\delta)$, where δ is taken from the interval $0.5 \le \delta \le 2$ [37] [38] [39] [46]. In part of works [40] [41] instead of classical intervals, interval fuzzy numbers are used.

In some works, in addition to triangles, other forms of functions, trapezoidal, gauge curves and the like are used. Unlike the previous one, there are different "softer" approaches in the Saaty scale fuzzification, where the confidence interval of the fuzzy number is defined after the completed pairing. Such an approach is shown in Kahraman et al. [42] and Pamucar et al. [43]. The fuzzification shown in these works takes into account the degree of uncertainty (β). Based on this, the new fuzzification, also used in this paper, is shown in Bozanic et al. [44], Pamucar et al. [45], Bozanic et al. [46] and Bojanic et al. [47]. The starting elements of this fuzzification are: 1) the introduction of fuzzy numbers instead of the classical numbers of the Saaty scale, 2) the introduction of the degree of conviction of decision makers/analysts/experts (DM/A/E) in the statements they give when compared in pairs - γ. The degree of conviction γ is defined at the level of each pairing, Table 3.

Table 3
Fuzzification of Saaty's scale using the degree of confidence

Definition	Standard values	Fuzzy number	Fuzzy number inverse values
The same importance	1	(1, 1, 1)	(1, 1, 1)
Weak dominance	3	$(3\gamma_{ji},3,(2-\gamma_{ji})3)$	$(1/3\gamma_{ji},1/3,1/(2-\gamma_{ji})3)$
Strong dominance	5	$(5\gamma_{ji},5,(2-\gamma_{ji})5)$	$(1/5\gamma_{ji},1/5,1/(2-\gamma_{ji})5)$
Very strong dominance	7	$(7\gamma_{ji},7,(2-\gamma_{ji})7)$	$(1/7\gamma_{ji},1/7,1/(2-\gamma_{ji})7)$
Absolute dominance	9	$(9\gamma_{ji},9,(2-\gamma_{ji})9)$	$(1/9\gamma_{ji},1/9,1/(2-\gamma_{ji})9)$
In between values	2, 4, 6, 8	$(x\gamma_{ji},x,(2-\gamma_{ji})x)$ x=2,4,6,8	$(1/x\gamma_{ji},1/x,1/(2-\gamma_{ji})x)$ x=2,4,6,8

Left and right distributions differ from one comparison to another when different degrees of confidence are introduced. They are changed according to the expression:

$$T = (t_1,t_2,t_3) = \left\{ \begin{array}{lll} t_1 = \gamma t_2, & t_1 \le t_2, & t_1,t_2 \in [1/9,9] \\ t_2 = t_2, & & t_2 \in [1/9,9] \\ t_3 = (2-\gamma)t_2, & t_3 \le t_2, & t_2,t_3 \in [1/9,9] \end{array} \right\} \quad (6)$$

where the value t_2 represents the value of the linguistic expression from the classical Saaty scale, which in the fuzzy number has the maximum affiliation $t_2=1$.

This approach to fuzzification also requires an altered initial matrix of decision-making, that is, a matching matrix in pairs (and in cases where the weighting

coefficients of the criteria and the ranking of the alternatives are defined). The modified matrix would have a new element - the degree of DM / A / E conviction that they would define in addition to the performed comparison:

$$A = \begin{array}{c} \\ K_1 \\ K_2 \\ \vdots \\ K_n \end{array} \begin{array}{c} \begin{array}{cccc} K_1 & K_2 & \dots & K_n \end{array} \\ \begin{bmatrix} a_{11};\gamma_{11} & a_{12};\gamma_{12} & \dots & a_{1n};\gamma_{1n} \\ a_{21};\gamma_{21} & a_{22};\gamma_{22} & \dots & a_{2n};\gamma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1};\gamma_{n1} & a_{n2};\gamma_{n2} & \dots & a_{nn};\gamma_{nn} \end{bmatrix} \end{array} \tag{7}$$

where $\gamma_{ji}=\gamma_{ij}$. Coming to the final results implies the further application of the standard steps of the AHP method.

3.2 Z Number

The concept of Z-number was proposed by Zadeh [48]. Kang et al. [49] [50] elaborate in detail the application of Z-number in an uncertain environment. Krohling et al. [51] show the application of Z-number with phasified methods TOPSIS and TODIM. Sahrom and Dom [52] elaborate a hybrid model using the AHP-Z-number-DEA method, while Azadeh and Kokabi [53] use the DEA method with Z-number. Azadeh et al. [54] use Z-number with the AHP method. Yaakob and Gegov [55] Z-number uses the TOPSIS method. Salari et al. [56] elaborate a novel earned value management model using Z-number. Z-number represents an ordered pair of fuzzy numbers that appear as $Z=(\tilde{A}, \tilde{B})$ [48]. The first component, fuzzy number $\tilde{A}$, represents the fuzzy limit of a particular variable *X*, while the second component fuzzy number $\tilde{B}$ represents, the reliability of the first component ($\tilde{A}$). The appearance of the Z-number with triangular fuzzy numbers is shown in Figure 4 [48].

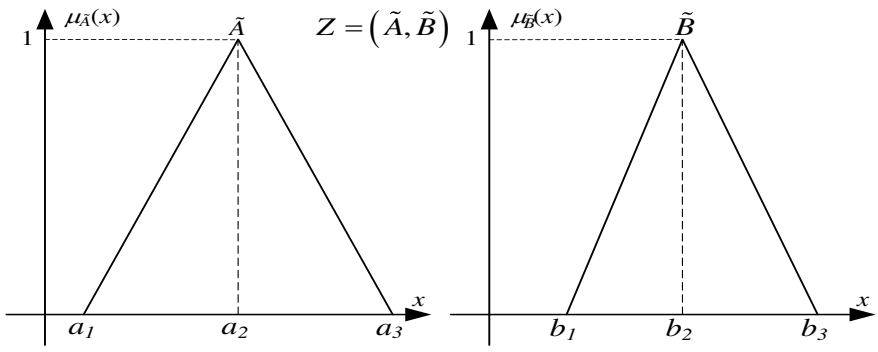

Figure 4
A-Simple Z-number [49]

The general record of triangular Z-numbers can be displayed as

$$\tilde{Z} = \left\{ \left(a_1, a_2, a_3; w_{\tilde{A}} \right), (b_1, b_2, b_3; w_{\tilde{B}}) \right\} \tag{8}$$

where the values $w_{\tilde{A}}$ i $w_{\tilde{B}}$ represent weight factors of fuzzy numbers $\tilde{A}$ referring to $\tilde{B}$, which for the initial Z-number the majority of authors defines as $w_{\tilde{A}} = w_{\tilde{B}} = 1$, $w_{\tilde{A}}, w_{\tilde{B}} \in [0,1]$ ($w_{\tilde{A}}$ is the height of the generalized fuzzy number and $0 \le w_{\tilde{A}} \le 1$) [57]. Defining values $w_{\tilde{B}} < 1$ would introduce a greater degree of uncertainty in defining Z-number When $w_{\tilde{A}} = w_{\tilde{B}} = 1$, then A and B is a normal fuzzy number; otherwise it is said to be a non-normal fuzzy number [58]. The transformation of the Z-number into the classical fuzzy number, with the presented evidence, is shown in Kang et al. [50]. This transformation consists of three steps:

Convert the second part ($\tilde{B}$) into a crisp number using the centered method [50]:

$$\alpha = \frac{\int x \mu_{\tilde{B}}(x) dx}{\int \mu_{\tilde{B}}(x) dx} \tag{9}$$

where $\int$ denotes an algebraic integration.

For triangular fuzzy numbers, the centroid is calculated by [59]:

$$\alpha = \frac{a_1 + a_2 + a_3}{3} \tag{10}$$

Add the weight of the second part ($\tilde{B}$) to the first part ($\tilde{A}$). The weighted Z-number can be denoted as [50]

$$\tilde{Z}^{\alpha} = \left\{ \langle x, \mu_{\tilde{A}^{\alpha}}(x) \rangle \middle| \mu_{\tilde{A}^{\alpha}}(x) = \alpha \mu_{\tilde{A}}(x) \right\} \tag{11}$$

which can be denoted by the Fig. 5a.

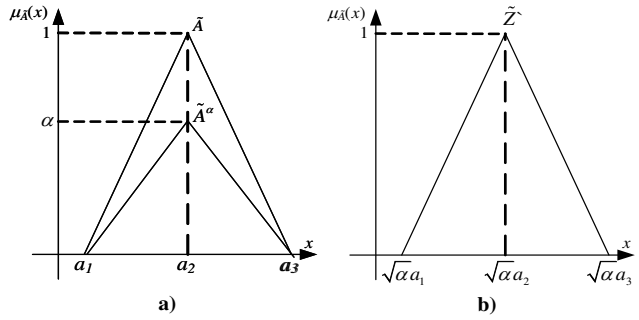

Figure 5
Z-number after multiplying the reliability (a) and the regular fuzzy number transformed from Z-number (b)

This can be written as [54]: $\tilde{Z}^{\alpha} = (a_1, a_2, a_3; \alpha)$

Convert the weighted Z-number into a regular fuzzy number. The regular fuzzy set can be denoted as [50]

$$\tilde{Z}' = \left\{ \langle x, \mu_{\tilde{Z}'}(x) \rangle \middle| \mu_{\tilde{Z}'}(x) = \mu_{\tilde{A}}\left(\frac{x}{\sqrt{\alpha}}\right) \right\} \tag{12}$$

$$\tilde{Z}' = \sqrt{\alpha} * \tilde{A} = (\sqrt{\alpha} * a_1, \sqrt{\alpha} * a_2, \sqrt{\alpha} * a_3) \tag{13}$$

and it can be present as Fig. 5b [50].

3.3 Fuzzy MABAC Method

The MABAC method is developed by Pamucar & Cirovic [60]. The basic setting of the MABAC method consists in defining the distance of the criteria function of every observed alternative from the border approximate area. The method was later used in several papers [44], [46], [47] [61]. The Fuzzy MABAC method consists of 6 steps [46]:

Step 1. Forming of the initial decision matrix ($\tilde{X}$). Matrix is formed with a grade of alternatives based on criteria $A_i = \left(\tilde{x}_{i1}, \tilde{x}_{i2} ..., \tilde{x}_{in}\right)$, where $\tilde{x}_{ij}$ represents the value of i-th alternative (i=1,2,...m), based on j-th criteria (j=1,2,...,n).

$$\tilde{X} = \begin{array}{c} \\ A_1 \\ A_2 \\ \dots \\ A_m \end{array} \begin{array}{c} \begin{array}{cccc} K_1 & K_2 & \dots & K_n \end{array} \\ \begin{bmatrix} \tilde{x}_{11} & \tilde{x}_{12} & \dots & \tilde{x}_{1n} \\ \tilde{x}_{11} & \tilde{x}_{22} & & \tilde{x}_{2n} \\ \dots & \dots & \dots & \dots \\ \tilde{x}_{1m} & \tilde{x}_{2m} & \dots & \tilde{x}_{mn} \end{bmatrix} \end{array} \tag{14}$$

Step 2. Normalization of the initial matrix elements ($\tilde{X}$).

The calculation of the elements of normalized matrix ($\tilde{N}$) depends on the type of criteria. For beneficial criteria this calculation is executed according to the expression:

$$\tilde{t}_{ij} = \frac{x_{ij} - x_i^-}{x_i^+ - x_i^-} \tag{15}$$

For detriment criteria the calculation is executed according to the expression:

$$\tilde{t}_{ij} = \frac{x_{ij} - x_i^+}{x_i^- - x_i^+} \tag{16}$$

Values x_{ij}, x_i^+, x_i^- represent elements of the initial matrix of decision-making ($\tilde{X}$). The values x_i^+, x_i^- are defined as explanied bellow

- $x_i^+ = \max(x_{1r}, x_{2r}, ..., x_{mr})$ - represent maximal values of the right distribution of fuzzy numbers of the observed criteria alternatives

- $x_i^- = \min(x_{1l}, x_{2l}, ..., x_{ml})$ - represent minimal values of the left distribution of fuzzy numbers of the observed criteria alternatives.

Consequently, the normalized matrix $(\tilde{N})$ is calculated

$$\tilde{N} = \begin{array}{c} \\ A_1 \\ A_2 \\ \dots \\ A_m \end{array} \begin{array}{c} \begin{array}{cccc} K_1 & K_2 & \dots & K_n \end{array} \\ \begin{bmatrix} \tilde{t}_{11} & \tilde{t}_{12} & \dots & \tilde{t}_{1n} \\ \tilde{t}_{11} & \tilde{t}_{22} & & \tilde{t}_{2n} \\ \dots & \dots & \dots & \dots \\ \tilde{t}_{1m} & \tilde{t}_{2m} & \dots & \tilde{t}_{mn} \end{bmatrix} \end{array} \tag{17}$$

Step 3. Calculation of the weighted matrix ($\tilde{\mathrm{V}}$) elements.

Elements of this matrix are calculated based on the following expression:

$$\tilde{v}_{ij} = w_i \tilde{t}_{ij} + w_i \tag{18}$$

In the previous expression $\tilde{t}_{ij}$ represents elements of the normalized matrix($\tilde{N}$), whereas w_i epresents weight coefficients of the criteria. Weighted matrix ($\tilde{\mathrm{V}}$) is visualized in the following way

$$\tilde{V} = \begin{bmatrix} \tilde{v}_{11} & \tilde{v}_{12} & \dots & \tilde{v}_{1n} \\ \tilde{v}_{21} & \tilde{v}_{22} & \dots & \tilde{v}_{2n} \\ \dots & \dots & \dots & \dots \\ \tilde{v}_{m1} & \tilde{v}_{m2} & \dots & \tilde{v}_{mn} \end{bmatrix} \tag{19}$$

Step 4. Determination of the approximate border area matrix ($\tilde{\mathrm{G}}$). The bordering approximate area for each criteria is determined based on the expression:

$$\tilde{g}_i = \left(\prod_{j=1}^{m} \tilde{v}_{ij} \right)^{1/m} \tag{20}$$

The matrix of approximate area $(\tilde{G})$ has a format n x 1, where n presents overall sum of criteria number and is represented in the following way

$$\tilde{G} = \begin{array}{c} \begin{array}{cccc} K_1 & K_2 & \dots & K_n \end{array} \\ \begin{bmatrix} \tilde{g}_1 & \tilde{g}_2 & \dots & \tilde{g}_n \end{bmatrix} \end{array} \tag{21}$$

Step 5. Calculation of the matrix elements of alternatives distance from the border approximate area ($\tilde{Q}$). The distance of alternatives from the border approximate area ($\tilde{q}_{ij}$) is defined with the expression:

$$\tilde{Q} = \tilde{V} - \tilde{G} \tag{22}$$

Afterwards the matrix is calculated $\tilde{Q}$

$$\tilde{Q} = \begin{bmatrix} \tilde{q}_{11} & \tilde{q}_{12} & \cdots & \tilde{q}_{1n} \\ \tilde{q}_{21} & \tilde{q}_{22} & & \tilde{q}_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \tilde{q}_{m1} & \tilde{q}_{m2} & \cdots & \tilde{q}_{mn} \end{bmatrix} \tag{23}$$

Step 6. Ranking of alternatives. The value estimation of criteria functions of alternatives is gained from the sum of the distance of alternatives from the border approximate areas ($\tilde{q}_i$). The ultimate values of criteria functions of alternatives are gained from the sum of elements of the matrix $\tilde{Q}$ in rows:

$$\tilde{S}_i = \sum_{j=1}^{n} \tilde{q}_{ij},\ j = 1, 2, ..., n,\ i = 1, 2, ..., m \tag{24}$$

Defuzzification of the extracted values $\tilde{S}_i$ yields the ultimate rang alternatives.

4 Application of the Model to the Ranking of Social Media Platforms

Based on previous experience and with regard to the fact that the surveyed experts are from the field of public relations, and available data from various sources, for example data from institute for statistics, relevant analytical websites, academic and expert literature which covers the field of social media for commercial purposes. The first experts summary of all data classified according to the criteria from available resources is presented in Table 4.

Table 4

Summary of all data classified according to the criteria

Alternative	Functionality	Target audience	Content richness	Purpose	Precence	Engagement
Blogs	5	100	4	5	0,1	3
Facebook	7	56	5	4	1721	5
Youtube	4	63	3	4	1325	3

Twitter	5	33	4	4	342	4
LinkedIn	3	12	4	4	380	2
Pinterest	5	8	3	3	73	4
Flicker	4	3	3	3	120	3
Instagram	7	63	3	3	500	4
Wikipedia	1	23	1	2	1500	1
Vimeo	4	3	3	4	40	3

The experts performed a comparison of the significance of the selected criteria, using the Saaty scale, and defined the degree of conviction in the comparisons they gave. This defined the initial decision matrices for each expert. Defining the degree of conviction was done using the fuzzy linguistic scale shown in Figure 6.

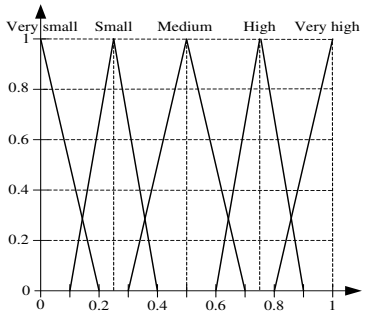

Figure 6

Fuzzy linguistic descriptors for evaluating the degree of conviction of experts

Table 4 shows the comparison of the criteria of the first expert, with degrees of confidence in data comparisons (degrees of conviction are shown in brackets). The implementation of the fuzzy AHP method will only be performed on the initial decision-making matrix of an expert.

Table 5

The initial matrix of deciding the expert 1 for defining the weight coefficients of the criterion

Criteria	C_1	C_2	C_3	C_4	C_5	C_6
C_1	1	3 (VH)	5 (H)	4 (H)	2 (H)	4 (VH)
C_2	1/3 (VH)	1	3 (H)	3 (VH)	1/3 (VH)	2 (S)
C_3	1/5 (H)	1/3 (H)	1	1/2 (VS)	1/6 (VH)	1/3 (M)
C_4	1/4 (H)	1/3 (VH)	2 (VS)	1	1/5 (M)	1/2 (VH)
C_5	1/2 (H)	3 (VH)	6 (VH)	5 (M)	1	6 (VH)
C_6	1/4 (VH)	1/2 (S)	3 (M)	2 (VH)	1/6 (VH)	1

$$\frac{\tilde{C}_1}{\tilde{C}_2} = \left(3*0.933, 3, \left(2-0.933\right)*3\right) = (2.8, 3, 3.2)$$

In the next step, the initial decision matrix was initialized using the terms given in Table 5. The Fuzzy comparison C1 / C2 was obtained as follows:

Where value γ_{ji}=0.933 represents the degree of conviction obtained using the expression (5), the displayed linguistic descriptors in Figure 6 and based on the expert's test that the degree of conviction in his claim is "very high". The given value was obtained as follows:

$$\gamma_{ji} = ((1-0.75)+(1-0.75))/3+0.75 = 0.933$$

All values of the fuzzified initial decision matrix are shown in Table 6. Next, the standard steps of the AHP method are applied. First, column addition is done. The results in the first column were done as follows:

$$\sum_{j=1}^{n} \tilde{C}_j^l = (1+0.31+0.16,+0.2+0.4+0.23,1+0.33+0.2+0.25+0.5+0.25,$$
$$1+0.36+0.27+0.33+0.67+0.27) = (2.3, 2.53, 2.9), n = 1, 2...6;$$

Table 6

Fuzzified initial decision matrix for expert 1

Criteria	C_1	C_2	C_3	C_4	C_5	C_6
C_1	(1,1,1)	(2.8,3, 3.2)	(3.75,5, 6.25)	(3,4,5)	(1.5,2, 2.5)	(3.73,4, 4.27)
C_2	(0.31,0.33, 0.36)	(1,1,1)	(2.25,3, 3.75)	(2.8,3, 3.2)	(0.31,0.33, 0.36)	(1,2, 3.5)
C_3	(0.16,0.2, 0.27)	(0.27,0.33, 0.44)	(1,1,1)	(0.26,0.5,1)	(0.16,0.17, 0.18)	(0.22,0.33, 0.67)
C_4	(0.2,0.25, 0.33)	(0.31,0.33, 0.36)	(1,2,3.87)	(1,1,1)	(0.13,0.2, 0.4)	(0.47,0.5, 0.54)
C_5	(0.4,0.5, 0.67)	(2.8,3, 3.2)	(5.6,6, 6.4)	(2.5,5, 7.5)	(1,1,1)	(5.6,6, 6.4)
C_6	(0.23,0.25, 0.27)	(0.29,0.5, 1)	(1.5,3,4.5)	(1.87,2,2.13)	(0.16,0.17, 0.18)	(1,1,1)
$\sum_{j=1}^{n} Cj$	(2.3,2.53, 2.9)	(7.47,8.16,9. 20)	(15.1,20, 25.77)	(11.43,15.5, 19.83)	(3.26,3.87, 4.62)	(13.02,13.83, 16.38)

Next, the fuzzy numbers were divided by a collection of columns, using standard fuzzy arithmetic:

$$\left(\frac{\tilde{C}_1}{\tilde{C}_2}\right) / \sum_{j=1}^{n} \tilde{C}_j = \left(\frac{1}{2.9}, \frac{1}{2.53}, \frac{1}{2.3}\right) = (0.35, 0.4, 0.43)$$

Then column addition is done:

$$\sum_{i=1}^{n} \tilde{C}_i^l = (0.35+0.3+0.15+0.15+0.33+0.23, 0.39+0.37+0.25+0.26+0.52+0.29,$$
$$0.43+0.43+0.41+0.44+0.77+0.35) = (1.5, 2.08, 2.84), n = 1, 2...6,$$

The other values are shown in Table 7.

Table 7
Values of AHP methods

Criteria	C_1	C_2	C_3	C_4	C_5	C_6	$\sum_{i=1}^{n} C_i$
C_1	(0.35,0.39, 0.43)	(0.30,0.37, 0.43)	(0.15,0.25, 0.41)	(0.15,0.26, 0.44)	(0.33,0.52, 0.77)	(0.23,0.29, 0.35)	(1.50,2.08, 2.84)
C_2	(0.11,0.13, 0.15)	(0.11,0.12, 0.13)	(0.09,0.15, 0.25)	(0.14,0.19, 0.28)	(0.07,0.09, 0.11)	(0.06,0.14, 0.29)	(0.57,0.83, 1.22)
C_3	(0.06,0.08, 0.12)	(0.03,0.04, 0.06)	(0.04,0.05, 0.07)	(0.01,0.03, 0.09)	(0.03,0.04, 0.05)	(0.01,0.02, 0.06)	(0.18,0.27, 0.44)
C_4	(0.07,0.1, 0.14)	(0.03,0.04, 0.05)	(0.04,0.1, 0.26)	(0.05,0.06, 0.09)	(0.03,0.05, 0.12)	(0.03,0.04, 0.04)	(0.25,0.39, 0.7)
C_5	(0.14,0.2, 0.29)	(0.30,0.37, 0.43)	(0.22,0.3, 0.42)	(0.13,0.32, 0.66)	(0.22,0.26, 0.31)	(0.34,0.43, 0.53)	(1.34,1.88, 2.64)
C_6	(0.08,0.1, 0.12)	(0.03,0.06, 0.13)	(0.06,0.15, 0.3)	(0.09,0.13, 0.19)	(0.03,0.04, 0.05)	(0.06,0.07, 0.08)	(0.36,0.55, 0.87)

The final values of the weight coefficients were obtained by dividing the sum of the rows with the number of criteria (n):

$$\sum_{i=1}^{n} \tilde{C}_i^1 / n = (\frac{1.5}{6}, \frac{2.08}{6}, \frac{2.84}{6}) = (0.25, 0.346, 0.473)$$

The steps shown are the standard steps of the AHP method, which resulted in the fuzzy weight coefficients of the criteria for the first expert ($\tilde{w}_{ej}$, j=1,...6, e=1,2,...11), Table 8.

Table 8
Fuzzy weight coefficients obtained using expert data 1

Criterion	$\tilde{w}_{ej}, e = 1$
C_1	(0.25, 0,346, 0.473)
C_2	(0.096,0.138,0.203)
C_3	(0.031,0.045,0.073)
C_4	(0.042,0.065,0.017)
C_5	(0.224,0.313,0.44)
C_6	(0.06, 0.092,0.145)

In order to compare the obtained results, the defuzzification of the obtained values was performed for the first expert and the normalization of value was done. The values of weight coefficients obtained from the comparison of the first expert are shown in Table 9.

Table 9
Comparative overview of weight coefficients obtained by classical and fuzzy AHP method

Criterion	Classical AHP method	Fuzzy AHP				
		Defuzzification $\tilde{w}_{1j}$ applying expression 4	Defuzzification $\tilde{w}_{1j}$ applying expression 5, λ=0	Defuzzification $\tilde{w}_{1j}$ applying expression 5, λ=0.5	Defuzzification $\tilde{w}_{1j}$ applying expression 5, λ=1	Rank
C_1	0.346	0.339	0.350	0.341	0.334	1
C_2	0.138	0.139	0.138	0.138	0.140	3
C_3	0.045	0.047	0.044	0.046	0.048	6
C_4	0.065	0.071	0.063	0.070	0.074	5
C_5	0.313	0.310	0.316	0.311	0.307	2
C_6	0.092	0.094	0.089	0.094	0.097	4

The differences observed between the application of classical AHP method and the fuzzy AHP methods are not large. Essential differences are noticed on the second and third decimals of weight coefficients, even though they represent minor differences. These differences also depend on the approach to fuzzification of the fuzzy weight coefficient of the criteria. The table provides two approaches or four budgets. As seen from the table, classical comparison in pairs is still dominant in the application of the method, while the degree of conviction is only an ancillary. These differences increase when multiple researchers use the analysis.

In further work, we calculated the weight coefficients for each expert. The obtained results were aggregated so one fuzzy number was obtained, which describes the weight coefficients, and which by using expression (5) is defuzzified. The final values of weight coefficients are given in Table 10.

Table 10
Final weight coefficients of the criterion

Criterion	Final weight coefficient (w)	Rank
C_1	0.351	1
C_2	0.113	3
C_3	0.063	6
C_4	0.074	5
C_5	0.318	2
C_6	0.081	4

After defining the weight coefficients, using the fuzzy MABAC method ranking of the alternative was performed. Ten alternatives are ranked. The values for the criteria C1, C2 and C5 are precisely defined, while the values for the criteria C3,

C4 and C6 are defined with two fuzzy or Z numbers: $\tilde{A}$ - which represents the value of the fuzzy number and $\tilde{B}$ - which represents the degree of the conviction of the experts in the described values. The initial matrix of decision-making ($\tilde{X}$) is shown in Table 11.

Table 11
The starting matrix of decision-making in ranking alternatives

Alternative	Alternative index	C_1	C_2	C_3		C_4		C_5	C_6	
				$\tilde{A}$	$\tilde{B}$	$\tilde{A}$	$\tilde{B}$		$\tilde{A}$	$\tilde{B}$
Blogs	A_1	5	4	(45,50,55)	VM	(4.5,5,5)	M	0.1	(2.5,3,3.5)	M
Facebook	A_2	7	5	(85,90,95)	M	(3.5,4,4.5)	V	1721	(4.5,5,5)	M
Youtube	A_3	4	3	(90,95,100)	S	(3.5,4,4.5)	VV	1325	(2.5,3,3.5)	VV
Twitter	A_4	5	4	(28,33,38)	S	(3.5,4,4.5)	S	342	(3.5,4,4.5)	S
LinkedIn	A_5	3	3	(29,34,40)	VM	(3.5,4,4.5)	M	380	(1.5,2,2.5)	VM
Pinterest	A_6	5	3	(22,27,32)	VV	(2.5,3,3.5)	VV	73	(3.5,4,4.5)	V
Flicker	A_7	4	3	(7,12,17)	V	(2.5,3,3.5)	S	120	(2.5,3,3.5)	S
Instagram	A_8	7	3	(81,86,91)	S	(2.5,3,3.5)	VM	500	(4.5,5,5)	VM
Wikipedia	A_9	1	1	(51,56,61)	VV	(1.5,2,2.5)	VM	1500	(1,1,1)	V
Vimeo	A_{10}	4	3	(0,4,9)	M	(3.5,4,4.5)	V	40	(2.5,3,3.5)	VV

In the next step, the phasing of linguistic expressions was performed using the scale shown in Figure 6. The display of the obtained values is given in Table 12.

Table 12
Fuzzified initial decision matrix by application of Z numbers in the ranking of alternatives

Alternative	Alter-native index	C_1	C_2	C_3		C_4		C_5	C_6	
				$\tilde{A}$	$\tilde{B}$	$\tilde{A}$	$\tilde{B}$		$\tilde{A}$	$\tilde{B}$
Blogs	A_1	5	4	(45,50,55)	0.067	(4.5,5,5)	0.25	0.1	(2.5,3,3.5)	0.25
Facebook	A_2	7	5	(85,90,95)	0.25	(3.5,4,4.5)	0.75	1721	(4.5,5,5)	0.25
Youtube	A_3	4	3	(90,95,100)	0.5	(3.5,4,4.5)	0.933	1325	(2.5,3,3.5)	0.933
Twitter	A_4	5	4	(28,33,38)	0.5	(3.5,4,4.5)	0.5	342	(3.5,4,4.5)	0.5
LinkedIn	A_5	3	3	(29,34,40)	0.067	(3.5,4,4.5)	0.25	380	(1.5,2,2.5)	0.067
Pinterest	A_6	5	3	(22,27,32)	0.933	(2.5,3,3.5)	0.933	73	(3.5,4,4.5)	0.75
Flicker	A_7	4	3	(7,12,17)	0.75	(2.5,3,3.5)	0.5	120	(2.5,3,3.5)	0.5
Instagram	A_8	7	3	(81,86,91)	0.5	(2.5,3,3.5)	0.067	500	(4.5,5,5)	0.067
Wikipedia	A_9	1	1	(51,56,61)	0.933	(1.5,2,2.5)	0.067	1500	(1,1,1)	0.75
Vimeo	A_{10}	4	3	(0,4,9)	0.25	(3.5,4,4.5)	0.75	40	(2.5,3,3.5)	0.933

Using the expression for transferring Z-numbers into standard fuzzy numbers, a new fazified initial matrix of decision-making is obtained, Table 13.

Table 13
Fuzzified initial matrix of decision making using classic fuzzy numbers

Alternative	Alternative index	C_1	C_2	C_3	C_4	C_5	C_6
Blogs	A_1	5	4	(0.904,1.033, 1.162)	(2.25,2.5, 2.5)	0.1	(1.25,1.5, 1.75)
Facebook	A_2	7	5	(2.25,2.5, 2.5)	(3.031,3.464, 3.897)	1721	(2.25,2.5, 2.5)
Youtube	A_3	4	3	(1.768,2.121, 2.475)	(3.381,3.864, 4.347)	1325	(2.415,2.898, 3.381)
Twitter	A_4	5	4	(2.475,2.828, 3.182)	(2.475,2.828, 3.182)	342	(2.475,2.828, 3.182)
LinkedIn	A_5	3	3	(0.904,1.033, 1.162)	(1.75,2,2.25)	380	(0.387,0.516, 0.645)
Pinterest	A_6	5	3	(2.415,2.898, 3.381)	(2.415,2.898, 3.381)	73	(3.031,3.464, 3.897)
Flicker	A_7	4	3	(2.165,2.598, 3.031)	(1.768,2.121, 2.475)	120	(1.768,2.121, 2.475)
Instagram	A_8	7	3	(1.768,2.121, 2.475)	(0.645,0.775, 0.904)	500	(0.904,1.033, 1.162)
Wikipedia	A_9	1	1	(0.966,0.966, 0.387)	(0.387,0.516, 0.645)	1500	(0.866,0.866, 0.866)
Vimeo	A_{10}	4	3	(1.25,1.5,1.75)	(3.031,3.464, 3.897)	40	(2.415,2.898, 3.381

After defining the initial decision matrix with standard fuzzy numbers, the application of the fuzzy MABAC method is performed. The final rank of the alternative is shown in Table 14.

Table14
Final rang of alternatives

Alternative	Altern. index.	Defuzzification $\tilde{w}_{1j}$ applying expression 4	Defuzzification $\tilde{w}_{1j}$ applying expression 5, λ=0	Defuzzification $\tilde{w}_{1j}$ applying expression 5, λ=0.5	Defuzzification $\tilde{w}_{1j}$ applying expression 5, λ=1	Rank
Blogs	A_1	-0.063	-0.072	-0.063	-0.055	6
Facebook	A_2	0.476	0.467	0.477	0.486	1
Youtube	A_3	0.182	0.164	0.181	0.199	2
Twitter	A_4	0.084	0.069	0.084	0.098	4
LinkedIn	A_5	-0.169	-0.177	-0.169	-0.161	10
Pinterest	A_6	0.024	0.005	0.023	0.042	5
Flicker	A_7	-0.079	-0.095	-0.080	-0.064	8

Instagram	A_8	0.104	0.094	0.104	0.114	3
Wikipedia	A_9	-0.157	-0.161	-0.157	-0.153	9
Vimeo	A_{10}	-0.079	-0.095	-0.079	-0.064	7

Regardless of which model of defadification is taken, the ranking alternative has not changed. However, a deeper analysis suggests that there are differences from one to another defadification, and that under certain circumstances the ranking of the alternative could be altered, primarily when it comes to the close values of the output parameters. The ranking of the first four alternatives were approved by the experts, who were interviewed independently of the application of the method. Matching expert views suggest that the model accurately depicts real world. On the other hand, the significance of the model is reflected in the ranking of other alternatives, as for other alternatives experts had very different views.

An unavoidable step in modeling is the assessment of its sensitivity. The sensitivity assessment was done by changing the weight coefficients of the criteria, using different scenarios, wherein each scenario the second criterion was favorable [62]. The display of weight coefficients according to the scenarios is given in Table 15.

Table 15

Weight coefficient in a different scenario

Criterion	S_0	S_1	S_2	S_3	S_4	S_5	S_6
C_1	0.351	0.5	0.1	0.1	0.1	0.1	0.1
C_2	0.113	0.1	0.5	0.1	0.1	0.1	0.1
C_3	0.063	0.1	0.1	0.5	0.1	0.1	0.1
C_4	0.074	0.1	0.1	0.1	0.5	0.1	0.1
C_5	0.318	0.1	0.1	0.1	0.1	0.5	0.1
C_6	0.081	0.1	0.1	0.1	0.1	0.1	0.5

The values obtained by applying different scenarios are given in Table 16.

Table 16

Ranking of alternatives by applying different scenarios

Altern. index	S_0		S_1		S_2		S_3		S_4		S_5		S_6	
	W	ra-nk	W	ra-nk	W	ra-nk	W	ra-nk	W	ra-nk	W	ra-nk	W	ra-nk
A_1	-0.063	6	0.002	6	0.046	5	-0.178	8	-0.034	7	-0.164	10	-0.094	7
A_2	0.476	1	0.402	1	0.413	1	0.312	1	0.339	1	0.503	1	0.277	1
A_3	0.182	2	0.091	5	0.102	3	0.154	4	0.268	2	0.300	2	0.221	3
A_4	0.084	4	0.143	3	0.187	2	0.253	2	0.149	3	0.056	4	0.198	4
A_5	-0.169	10	-0.206	9	-0.128	9	-0.253	9	-0.150	8	-0.150	9	-0.281	9
A_6	0.024	5	0.125	4	0.069	4	0.246	3	0.138	4	-0.024	6	0.252	2

A_7	-0.079	8	-0.026	8	-0.015	7	0.114	5	-0.025	6	-0.097	7	0.015	6
A_8	0.104	3	0.162	2	-0.027	8	0.025	6	-0.173	9	-0.021	5	-0.121	8
A_9	-0.157	9	-0.388	10	-0.377	10	-0.312	10	-0.349	10	0.061	3	-0.290	10
A_{10}	-0.079	7	-0.019	7	-0.008	6	-0.056	7	0.118	5	-0.109	8	0.111	5

By analyzing the obtained results, it can be determined that there are differences in ranks when applying different scenarios, which indicates that the model is sensitive to changes. It is noted that alternative A2 is always ranked as the first one, which can indicate the insensitivity of the model, but the specific values of the output results are in favor of the sensitivity of the system. In support of this, the results obtained with scenario number 6, where the values of the first three ranged alternatives are very close. A constant rank of Alternative A2 indicates the dominance of the input parameters of this alternative relative to others by all criteria. The highest ranking range is present in the alternative A8, where the ranges go from 2 to 9, and alternatives A9 where the ranges go from 3 to 10. In other alternatives, this ranking range is much lower. All of the above indicates that the model can tolerate even minor errors in defining the weight coefficients of the criteria.

For the mathematical determination of the correlation of ranks, the values of Spirman's coefficient were used:

$$S = 1 - \frac{6\sum_{i=1}^{n} D_i^2}{n(n^2 - 1)} \tag{11}$$

where is:

S - the value of the Spirman coefficient,

Di - the difference in the rank of the given element in the vector w and the rank of the correspondent element in the reference vector,

n - the number of ranked elements.

The rank of each criterion - the alternative is determined based on the weight coefficient vector $w=(w_1, w_2, ..., w_n)$.

Spirman's coefficient takes values from the interval [-1,1]. When the ranks of the elements completely coincide, the Spirman coefficient is 1 ("ideal positive correlation"). When the ranks are completely opposite, the Spirman coefficient is -1 ("ideal negative correlation"), that is, when S = 0 the ranks are unregulated.

Table 17

Spirman's coefficient values

	S_0	S_1	S_2	S_3	S_4	S_5	S_6
S_0	-	0.915	0.782	0.782	0.685	0.636	0.721
S_1	-	-	0.733	0.806	0.588	0.455	0.673

S_2	-	-	-	0.879	0.939	0.418	0.915
S_3	-	-	-	-	0.867	0.539	0.903
S_4	-	-	-	-	-	0.455	0.952
S_5	-	-	-	-	-	-	0.430
S_6	-	-	-	-	-	-	-

As seen from the table of Spirman's coefficient values, they are very high in a large number of cases. It can be said that there are no unordered ranks, which leads to the same conclusion as after the analysis of the previous picture, i.e., that the model can tolerate even minor errors in defining the weight coefficients of the criteria.

Conclusion

The Ministry of Defense of the Republic of Serbia uses a criteria selection and ranking methodology in choosing the social media platforms it uses in communication with the public. Statistical indicators and analyses of the usage of social media by the public administration of the Republic of Serbia, which are conducted annually by the Office for Electronic Administration of the Republic of Serbia and the Faculty of Organizational Sciences of the University of Belgrade [63], identify the Ministry of Defense as the most efficient governmental institution in the application of social media.

Further development of social media ranking methodology can be articulated in two directions: Firstly, the development of an application using set criteria and weight factors for automatic ranking according to the described methodology, thus speeding up data calculation and giving the experts the opportunity to correct errors in criteria and weight factors assessment. Secondly, the selection process and the number of criteria affecting ranking can be further improved in direct dependence on future development and usage of social media on a global level.

Literature

[1] K. H. Yoo, R. J. Kim: Review How U. S. state tourism offices use online newsrooms and social media in media relations, Public Relations Review, Vol. 39, 2013, pp. 534-541

[2] R. Medaglia, L. Zheng: Mapping government social media research and moving it forward: A framework and a research agenda, Government Information Quarterly, 2017, Vol. 34, pp. 496-510

[3] J. Kietzmann, et al.: Social media? Get serious! Understanding the functional building blocks of social media, Business Horizons, 2011, Vol. 54, pp. 241-251

[4] T. M. Nisar, G. Prabhakar, L. Strakova: Social media information benefits, knowledge management and smart organizations, Journal of Business Research, 2019, Vol. 94, pp. 264-272

[5] T. Grizane, I. Jurgelane: Social media impact on Bussiness Evaluation, Procedia Computer Science, 2017, Vol. 104, pp. 190-196

[6] J. I. Criado, J. Villodre: Public employees in social media communities: Exploring factors for internal collaboration using social network analysis, First Monday, 2018, Vol. 23, No. 4

[7] G. F. Khan, H. Y. Yoon, H. W. Park: "Social media use in the public sector: a comparative study of the Korean & US Government", The 8th International Conference on Webometrics, Informatics and Scientometrics & 13th COLLNET Meeting, 2012, Seoul

[8] R. S. Almazan, J. R. Gil-Garcia: Are government internet portals evolving towards more interaction, participation, and collaboration? Revisiting the rhetoric of e-government among municipalities, Government Information Quarterly, 2012, Vol. 29, pp. 72-81

[9] A. Zerfass, D. M. Schramm: Review Social Media Newsrooms in public relations: A conceptual framework and corporate practices in three countries, Public Relation Review, 2014, Vol. 40, pp. 79-91

[10] A. J. Meijer, R. Tolenvlied: Social Media and the New Organization of Government Communications: An Empirical Analysis of Twitter Usage by the Dutch Police, American Review of Public Administration, 2016, Vol. 46, pp. 143-161

[11] R. Dilenschneider: The AMA Handbook of Public Relations, Amacom, 2010

[12] C. Knox: Public Administrators Use of Social Media Platforms: Overcoming the Legitimacy Dilemma, Administration and Society, 2014, Vol. 48, pp. 477-496

[13] M. Feeney, E Welch: Technology – Task Coupling: Exploring Social Media Use and Managerial Perceptions of E-government, American Review of Public Administration, 2016, Vol. 46, pp. 162-179

[14] J. Kietzmann, et al.: Social media? Get serious! Understanding the functional building blocks of social media, Business Horizons, 2011, Vol. 54, pp. 241-251

[15] https://www.alexa.com/topsites/countries, Date: 30.06.2019

[16] A. Kaplan, M. Haenlein, Users of the world, unite! The challenges and opportunities of Social Media, Business Horizons, 2010, 53, pp. 59-68

[17] R. L. Daft, R. H. Lengel: Information Richness: A New Approach to Managerial Behaviour and Organizational Design. In Research in Organizational Behaviour, CT JAI Press, 1984

[18] A. Drobnjak, Fuzzy Social Media Clustering, Master thesis, University of Fribourg, Switzerland, 2012

[19] K. Djuric-Atanasievski, Z. Bobar: Social Media and Public Sector – The experience of the use of Social Media in State Organizations and Serbian Defense Structure, Symorg, Zlatibor, 2016, Vol. 15, pp. 125-133

[20] M. Cho, T. Schweickart, T. Haase: Public engagement with nonprofit organizations on Facebook, Public Relation Review, 2014, Vol. 40, pp. 565-567

[21] T. L. Saaty: The analytic hierarchy process, McGraw-Hill, New York, 1980

[22] L. A. Zadeh: A Rationale for Fuzzy Control, Journal of Dynamic Systems, Measurement and Control, 1972, Vol. 94, pp. 3-4

[23] L. A. Zadeh: Outline of a new approach to the analysis of complex systems and decision processes, IEEE Transactions on Systems, Man and Cybernetics, 1973, SMC-3(1), pp. 28-44

[24] L. Zadeh: Fuzzy sets, Information and control, 1965, Vol. 8, pp. 338-353

[25] D. Pamucar, D. Bozanic, A. Milic: „Selection of a course of action by Obstacle Employment Group based on a fuzzy logic system“, YUJOR, 2016, Vol. 26, pp. 75-90, doi: 10.2298/YJOR140211018P

[26] K. Husain, M. Rashid, N. Vitkovic, J. Mitic, J. Milovanovic, M. Stojkovic: Geometrical models of mandible fracture and plate implant, Facta Universitatis, Series: Mechanical Engineering, 2018, 16(2), pp. 369-379

[27] L. M. Seiford: The evolution of the state-of-art (1978-1995), Journal of Productivity Analysis, 1996, Vol. 7, pp. 99-137

[28] T. S. Liou, M. J. Wang: Ranking fuzzy numbers with integral value. Fuzzy Sets and Systems, 1992, Vol. 50, pp. 247-256

[29] M. Milicevic: The Expert Evaluation, MC „Defense“, Belgrade, 2014

[30] D. Bozanic, D. Pamucar, D. Bojanic: „Modification of the Analytic Hierarchy Proces (AHP) Method using fuzzy logic: fuzzy AHP approach as a support to the decision making process concerning engagement of the Group for Additional Hindering“, Serbian Journal of Management, 2015, Vol. 10, pp. 151-171

[31] D. Pamucar, D. Bozanic, V. Lukovac, N. Komazec: Normalized weighted geometric bonferroni mean operator of interval rough numbers – application in interval rough DEMATEL-COPRAS. Facta Universitatis, Series: Mechanical Engineering, 2018, Vol. 16, pp. 171-191

[32] F. T. Bozbura, A. Beskese, C. Kahraman: Prioritization of human capital measurement indicators using fuzzy AHP, Expert Systems with Applications, 2007, Vol. 32, pp. 1100-1112

[33] M. T. Isaai, A. Kanani, M. Tootoonchi, H. R. Afzali: Intelligent timetable evaluation using fuzzy AHP, Expert Systems with Applications, 2011, Vol. 38, pp. 3718-3723

[34] K. Mandic, B. Delibasic, S. Knezevic, S. Benkovic: Analysis of the financial parameters of Serbian banks through the application of the fuzzy AHP and TOPSIS methods. Economic Modelling, 2014, Vol. 43, pp. 30-37

[35] F. R. L. Junior, L. Osiro, L. C. R. Carpinetti: A comparison between Fuzzy AHP and Fuzzy TOPSIS methods to supplier selection, Applied Soft Computing, 2014, Vol. 21, pp. 194-209

[36] A. Ruiz-Padillo, A. J. Torija, A. F. Ramos-Ridao, D. P. Ruiz: Application of the fuzzy analytic hierarchy process in multi-criteria decision in noise action plans: Prioritizing road stretches, Environmental Modelling & Software, 2016, Vol. 81, pp. 45-55

[37] B. Srdjevic, Y. Dantas, P. Modeiros: Fuzzy AHP Assessment of Water Management Plans, Water Resour Manage, 2008, Vol. 22, pp. 877-894

[38] V. Nikolic, M. Milovancevic, D. Petkovic, D. Jocic, M. Savic: Parameters forecasting of laser welding by the artificial intelligence techniques, Facta Universitatis, Series: Mechanical Engineering, 2018, Vol. 16, pp. 193-201

[39] D. Pamučar, V. Lukovac, D. Božanić, N. Komazec: Multi-criteria FUCOM-MAIRCA model for the evaluation of level crossings: case study in the Republic of Serbia, Operational Research in Engineering Sciences: Theory and Applications, 2018, Vol. 1, pp. 108-129

[40] L. Abdullah, L. Najib: A new type-2 fuzzy set of linguistic variables for the fuzzy analytic hierarchy process. Expert Systems with Applications, 2014, Vol. 41, pp. 3297-3305

[41] C. Kahraman, B. Öztayşi, I.U. Sari, E. Turanoğu: Fuzzy analytic hierarchy process with interval type-2 fuzzy sets. Knowledge-Based Systems, 2014, Vol. 59, pp. 48-57

[42] D. Pamucar, G. Cirovic, D. Sekulovic: Development of an integrated transport system in distribution centers: A FA'WOT analysis, Technical Gazzete, 2015, Vol. 22, pp. 649-658

[43] D. Bozanic, D. Pamucar, D. Bojanic: „Modification of the Analytic Hierarchy Proces (AHP) Method using fuzzy logic: fuzzy AHP approach as a support to the decision making process concerning engagement of the Group for Additional Hindering“, Serbian Journal of Management, 2015, Vol. 10, pp. 151-171

[44] D. Bozanic, D. Pamucar, S. Karovic: Use of the fuzzy AHP - MABAC hybrid model in ranking potential locations for preparing laying-up positions, Military Technical Courier, 2016, Vol. 64, pp. 705-729, doi: 10.5937/vojtehg64-9261

[45] D. Pamucar, D. Bozanic, D. Kurtov: Fuzzification of the Saaty's scale and a presentation of the hybrid fuzzy AHP-TOPSIS model: an example of the selection of a brigade artillery group firing position in a defensive

operation, Military Technical Courier, 2016, Vol. 64 (4), pp. 966-986, doi: 10.5937/vojtehg64-9262

[46] D. Bozanic, D. Tesic, J. Milicevic: A hybrid fuzzy AHP-MABAC model: Application in the Serbian Army – the selection of the location for deep wading as a technique of crossing the river by tanks, Decision Making: Applications in Management and Engineering, 2018, Vol. 1, pp. 143-164

[47] D. Bojanic, M. Kovac, M. Bojanic, V. Ristic: Multi-criteria decision-making in A defensive operation of THE Guided anti-tank missile battery: an example of the hybrid model fuzzy AHP – MABAC, Decision Making: Applications in Management and Engineering, 2018, Vol. 1, pp. 51-66

[48] L. A. Zadeh: A note on Z-number, Information Sciences, 2011, Vol. 181, pp. 2923-2932

[49] B. Kang, D. Wei, Y. Li, Y. Deng: Decision Making Using Z-numbers under Uncertain Environment, Journal of Computational Information Systems, 2012, Vol. 8, pp. 2807-2814

[50] B. Kang, D. Wei, Y. Li, Y. Deng: A Method of Converting Z-number to Classical Fuzzy Number, Journal of Information & Computational Science, 2012, Vol. 9 (3), pp. 703-709

[51] R. A. Krohling, G. Artem dos Santos, A. G. C. Pacheco: TODIM and TOPSIS with Z-number, arxiv.org/pdf/1609.05705

[52] N. A. Sahrom, R. M. Dom: A Z-number extension of the hybrid Analytic Hierarchy Process-Fuzzy Data Envelopment Analysis for risk assessment. In ICREM7 - Proceedings of the 7th International Conference on Research and Education in Mathematics: Empowering Mathematical Sciences through Research and Education, Kuala Lumpur, Malaysia, 2015, pp. 19-24, doi: 10.1109/ICREM.2015.7357019

[53] A. Azadeh, R. Kokabi: Z-number DEA: A new possibilistic DEA in the context of Z-numbers, Advanced Engineering Informatics, 2016, Vol. 30, pp. 604-617

[54] A. Azadeh, M. Saberi, P. Pazhoheshfar: Z-AHP: A Z-number Extension of Fuzzy Analytical Hierarchy Process (2013) 7th IEEE International Conference on Digital Ecosystems and Technologies (DEST), pp. 141-147

[55] A. M. Yaakob, A. Gegov: Fuzzy Rule-Based Approach with Z-Numbers for Selection of Alternatives using TOPSIS, IEEE International Conference on Fuzzy Systems At Istanbul, Turkey, 2015, pp. 1-8

[56] M. Ciavarella, G. Carbone, V. Vinogradov: A critical assessment of Kassapoglou's statistical model for composites fatigue, Facta Universitatis, Series: Mechanical Engineering, 2018, Vol. 16, 115-126

[57] R. Chutia, S. Mahanta, D. Datta: Linear equations of generalized triangular fuzzy numbers, Annals of Fuzzy Mathematics and Informatics, 2013, Vol. 6 (2), pp. 371-376

[58] R. Saneifard: On The Ranking Fuzzy Numbers Using Signal/Noise Ratios, Int. J. Industrial Mathematics, 2015, Vol. 7 (1), pp. 213-219

[59] H. J. Zimmermann: Fuzzy Set Theory and It's Applications, Second Revised ed., Kluwer Academic Publishers, Boston/Dordrecht/London, 1991

[60] D. Pamucar, D. Cirovic: The selection of transport and handling resources in logistics centers using Multi-Attributive Border Approximation area Comparison (MABAC), Expert systems with applications, 2015, Vol. 42(6), pp. 3016-3028

[61] Z. Nunić: Evaluation and selection of the PVC carpentry manufacturer using the FUCOM-MABAC model, Operational Research in Engineering Sciences: Theory and Applications, 2018, Vol. 1 (1), pp. 13-28

[62] D. Pamucar, D. Bozanic, A. Randjelovic: Multi-criteria decision making: An example of sensitivity analysis, Serbian Journal of Management, 2017, Vol. 12(1), pp. 1-27, DOI:10.5937/sjm12-9464

[63] [http://www.deu.gov.rs/latinica/dokumenti.php Date: 15.08.2018

Comparative Study on Plant Type Specific LED Light Source Design Parameters

Csaba Kárász*, József Kopják

Óbuda University, Kandó Kálmán Faculty of Electrical Engineering, Bécsi út 96/b, H-1034 Budapest, Hungary, karasz.csaba@cl.uni-obuda.hu*, kopjak.jozsef@kvk.uni-obuda.hu

Abstract: The aim of this study is to highlight the correlations among the main parameters of lighting systems, for the agricultural industry. With these data, engineers could enhance the efficiency of their light source, in regards to different plant types. Statistical analysis has been conducted on the results of several experiments to find out if there is any correlation among different LED driver settings, which could affect light source design. The experiments have been re-grouped according to plant types and used variables. Two major groups have been examined: the experiments made with lettuce and with microalgae (in which most studies were made in the topic of photo-bioreactor enhancement). Smaller, sub-groups, have been clustered together with similar test variables as well. Significant correlations have been found between the variables of: light intensity (photosynthetic photon flux density) and photoperiod, light intensity and frequency, and light intensity and light spectrum as well. The results are in congruence with other studies and the summary of the analysis could be used to optimize light source design.

Keywords: Indoor Plant Cultivation; PPFD; Photosynthesis; Pulse Width Modulation, LED light source driver

1 Introduction

Production efficiency is more and more crucial on the field of agriculture. Sunlight is dependent on weather conditions, time of the season and other time and place dependent parameters. Therefore, artificial light has to be provided to enhance or substitute natural light in plant production.

This paper highlights the relationship between pulse width modulated (PWM) LED light source system parameters such as wavelength, frequency, duty cycle and light intensity (PPFD). This information provides support for engineers who are involved in plant factory design process.

Different types of light sources are used to enhance plant growth in agriculture, but light emitting diode (LED) could be the optimal choice. Its response time is

quite fast, has much longer lifetime, very good controllability and has very good efficiency. Also, LEDs radiate electromagnetic waves in a quite narrow wave band. This attribute could give a perfect solution to "assemble" the spectrum necessary for most applications [38].

Different plants might need different light source settings. If the correlation among different light source parameters could be known plant specifically, the control of light sources could be optimized. It would be possible to use less electricity and to achieve better plant quality and quantity. The effect of change of different parameters of light source on plant growth have been examined by previous researches. The following variables were tested in previous studies: duty cycle, frequency, wavelength, light intensity, photoperiod and phase difference of different wavelengths of a light source. However, only few of these parameters have been examined in the same experiment. Jao et al. [21] [22] tested different combinations of duty cycle, frequency and photoperiod, while Mori et al. [28] and Park et al. [34] checked frequency and duty cycle (Table 3). Lot of experiments made to find out the results of frequency change (Kanechi, M. [25], Son, K.-H. et al., [45], Jannsen M.G.J. [20], Vejrazka, C. et al., [49]). Others checked wavelength (light color) difference (Pardo, G.P. et al. [33], Yago, T. et al. [52] [53]) and the phase difference between different wavelengths of light (Shimada, A. et al. [44]) or the combination of them (Senol, R. et al. [40] [41]). Timothy, M. et al. [48] checked the effect of light intensity change (via the help of the change of duty cycle). Berkovich, Y. A. et al. [6] summarized the results of studies and experiments.

Usually, only a few parameters are tested in a research. In this paper a statistical analysis is introduced to be used on already performed experiment data to synthesize their best results. It would be easier to optimize the parameters of PWM driven light sources and save energy cost with the found correlations. Based on this information, depending on the aim of a cultivation facility (which plant types to grow), it is possible to enhance the feasibility of a plantation.

In the plant photosynthesis part, the light quality needs of plants are investigated. This information is vital to achieve perfect light conditions for plants and to be able to interpret data in a statistical analysis. It is possible to reach (or go beyond) the (photosynthetic) saturation levels (photoinhibition) [10] under optimal conditions. In the present paper the effects of electromagnetic radiation on plants are categorized according to PWM controlled LED light parameters. Part of the raw data is taken from a previous article [24] and used from an entirely new point of view. This way the analyzed parameters could be viewed from an engineering (PWM drive) point of view.

Photosynthetic photon flux density (PPFD), is the measure of light energy arriving to plants. In the PPFD parameter calculation section of the paper, it is highlighted how to convert light source parameters according to the processes of plant photosynthesis procedures.

The plant specific light source parameter part shows the results of the compared studies. It gives information about optimal LED driver settings (PWM duty cycle, frequency etc.) for specific plants.

The statistical analysis part introduces the used analysis methods to achieve a synthesis of the analyzed experiments.

In the last part of the paper, the results of the statistical analysis are highlighted and the findings of the study are summarized in the conclusion. This procedure could give support to the realization of a more efficient PWM driven LED light source design for plant cultivation.

2 The Photosynthesis of Plants

2.1 The Effect of Different Light Wavelengths on Plants

2.1.1 The Light Sensors of Plants – Photoreceptors

Photoreceptors in plants mostly not just sense light but absorb it as well – and according to their types, they absorb different band of light. These photoreceptors transmit - transduce signals to the crops to be able to change their structure.

Several photoreceptor signals work together, which result in plant response or in a morphological change of the plant. These are the phytochromes, cryptochromes, rhodopsins, xanthopsins and phototropin (light-oxygen-voltage (LOV) sensors) [29]. The newest discovered photoreceptor is the UV resistance locus 8, which absorbs UV-B light [6].

It is important to find a photoreceptor signal's effect and also have to find a way to stimulate the photoreceptor with light.

2.1.2 The Energy Harvesters of Plants - Photosynthetic Pigment Groups

In plants photoreceptors and the materials which are contained in their cells/structure are effected by light as well. The chloroplast has several parts absorbing electromagnetic waves and reacting to it. Only certain parts of the electromagnetic spectrum affect them and only certain parts of it are absorbed by them. As a consequence of the "color" sensitivity of the chloroplast, it has a certain color as well: therefore they can be called pigments [5]. They are responsible for different metabolic processes of plants. These are photosynthesis, thermal balance, etc..: the state of the plant and its behavior [12] [5].

Since pigments absorb light in certain wavebands only, it is possible to group them accordingly. Chlorophylls, Phycobilins and Carotenoids are the major pigment groups.

Chlorophyll is one of the major pigment group, consists of "Chlorophyll a" and "b", (c1, c2 and d). It consists of a phytol tail and a porphyrin ring, which respectively captures light. With its electrons activates the photosynthesis. "Chlorophyll b" absorbs a different band of light and helps "Chlorophyll a" with its energy. It reacts to green (450-475 [nm]) and orange-red (630-675 [nm]) light. Chlorophyll content increases when R : FR ratio increases – that is also the time when photosystem II to photosystem I decreases (PSII/PSI) [14].

Carotenoids have the task to deal with the excess of light. They protect other pigments (Chlorophylls at first place) and absorb light from the blue (400 [nm]) until the green (550 [nm]) band.

Phycobilins are present mostly in cyanobacteria and red microalgae. They replace the accessories of Chlorophyll in terms of light absorption accessory – antennae. They absorb light bands in the blue, green, yellow or orange (400-650 [nm]) parts [27].

Table 1
Photopigment sensitivity to different bands of electromagnetic stimulus. The effect of electromagnetic stimulus on plant morphology

Photopigment name	**Color Sensitivity**	**Area controlled / effected by the pigment**
Chlorophyll a, b (c1,c2,d)	green (450-475 nm) and orange-red (630-675nm)	Captures light, and with its electrons activates the photosynthesis. Chlorophyll b absorbs a different band of light and helps Chlorophyll a with its energy.
Phycobilin (primarily in cyanobacteria and red microalgae)	blue, green, yellow or orange (400-650 nm)	They replace the accessories of Chlorophyll in terms of light absorption accessory – antennae.
Carotenoid	from the blue (400 nm) until the green (550 nm)	It deals with the excess of light – protecting other pigments (Chlorophylls at first place).

Table 2
Photoreceptor sensitivity to different bands of an electromagnetic stimulus. The effect of electromagnetic stimulus on plant morphology

Photoreceptor name	Color Sensitivity	Area controlled / effected by the receptor
Phytochrome (might be complementer system with Cryptochrome) ABCF/E	Red : Far red ***ratio*** *(shade)*	Early flowering - stem elongation, Fatty acid regulation during seed maturation, Chlorophyll and carotenoid increase - Reduces during the End Of the Day
Cryptochrome (might be complementer system with Phytochrome) Cry1,Cry2	UV-A and Blue	Cell expansion reduction
Phototropin (light-oxygen-voltage (LOV) sensor) Phot1, Phot2	Blue	Chloroplast and leaf movement, stomatal opening, phototropism, onset of flowering, entrainment of the circadian clock, regulate aspects of morphogenesis, , de-etiolation
Rhodopsin (in Microalgae)	Light intensity (color independent)	Fototaxis direction indicator
UVR8 (UV res. Locus 8)	UV-B	Light dependent gene expression reg.

Taken into account the above mentioned information, the following wavebands are the most important: UV-B, UV-A, blue, red and far-red light. However, green [6] [11], yellow and the orange band also affects plants (Table 2).

2.2 The Speed of Different Processes in Photosynthesis

The oxidization of water (via a tetra manganese cluster [51]) is the base of the whole photosynthesis. It gives the base material (electrons and protons) for the Electron Transport Chain (ETC). This process is in the thylakoid membrane of the chloroplast. The ETC process could be called an energy harvester procedure since photons activate it. The performance of the ETC could be characterized with the

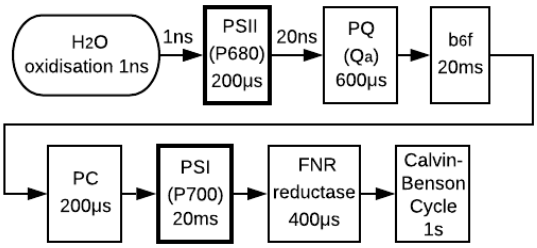

Figure 1
Process of the electron transport chain [25]

parameter called Linear Electron Flux (LEF) as well [5]:

From the above showed processes the PSII and PSI reaction (centers) are the light (induced) reactive phases. It is important to take into account the speed of these processes. The frequency of light reaching these centers are essential in terms of the efficiency of photosynthesis. The speed of the reaction center PSII is approximately 200 μs and the speed of PSI is 20 ms [25].

In parallel with these processes works the ATP synthase, but its reactions are non-photochemical. Plants create adenozyne-tri-phosphate (ATP) with the help of the light reaction of photosynthesis. This gives energy to the plants in and also nicotinamide – adenine-dinucleotide-phosphate ($NADP^+$). It reduces power to organic materials. These are synthesized with the help of Photosystem II (P680), Photosystem I (P700) and ATP synthase.

The Calvin(-Benson) cycle happens in the thylakoid lumen and it could be one of the end phases of the energy conversion of photosynthesis. During this reaction happens the carbon assimilation. The created ATP and $NADPH_2$ helps to assimilate CO_2 into organic materials (recreate cells, or store it) [5] [5]. This process is directly independent of light. It is dependent on the processes (end products) of PSII and PSI, it is also indirectly dependent on light.

Mostly the harmony of the ETC process makes it possible for pulsed light to give better results than continuous light. This happens with the re-oxidation of the plastoquinin pool in the chloroplast/photosynthetic membrane. Higher photon flux density (PFD) could be reached – and with that higher quantities could be reached in plant production [30]. It is important to have an ***abundant supply of*** other input materials than light as well (depending on plant type: CO_2, H_2O, N_2, P etc…) [43]. Moreover, it is important to state, that there are studies which conducted research with the concentrated and pulsed light of the sun. The results of these studies showed that plant cultivation with pulsed sunlight gives better results than continuous sunlight [13]. Also, other studies state that sunlight itself can not arrive without interruption to most plants. Different obstructions are in their way, such as leaves on trees [25].

2.3 Light Energy Required for Plants

It is very important to note that light energy input for plants is very much connected with the daily light period used. The total daily amount of photons could be the same with longer photoperiod. For example, with 16/24 instead of 12/12 would need lower PPFD (i.e.: with this, the photoinhibition could be avoided).

The most efficient power level could be different at each type of plant. According to different researches, the minimum level of PPFD is 0.1 [$\mu molm^{-2}s^{-1}$] which is a measurement threshold [5], but the efficient PPFD usage could reach 1500 [$\mu molm^{-2}s^{-1}$] as well [6]. However, common values are 200-300-400 [$\mu molm^{-2}s^{-1}$]). The best solution is to use measurements to confirm the best illumination input.

There are different ways to measure the level of photosynthesis in vivo. One indirect method is to measure the level of chlorophyll *a* fluorescence with the help of a (portable) fluorimeter [5]. Although this factor belongs to the PSII center, it is correlated with the efficiency of the whole photosynthesis. With the measurement of the minimum and maximum level of fluorescence, it is possible to find the best operating efficiency of PSII.

The best possible power settings could be skewed by the indirect effects of other parts of the ETC on PSII. Such effects are the regeneration time, resource needs and assimilation rates of PSI center. Furthermore, ATP synthase and the Calvin Cycle (among others) could affect power settings. Furthermore, the effect of the environment of the plants have to be taken into account as well (possible abiotic stresses like gases present, temperature, humidity, nutrients) [9].

3 PPFD Parameter Calculation from Flux of LED

An LED is different from a diode in terms of its (emitting) nature. This emission is measured in luminous flux (photometric power [lm]) or radiant flux (radiometric value [W]). It is also characterized by the wavelength of the emitted light [nm].

In the UV and IR bands the radiance: the intensity of the emission is represented by the SI unit of radiometric flux: Φ_e; watt [W] or [mW] – which is a radiometric value.

Component catalogues use the SI unit of Luminous flux: Φ_v, lumen [lm] as the measurement unit for the emission values in the visible region of the electromagnetic wave spectrum. This is a photometric value, representing luminous energy (which has a visible – luminous – nature). This parameter is based on the photopic response of the human eye (relative spectral sensitivity function ($V(\lambda)$) [33]). The photoreceptors of the human visual perception differ from plants (in functionality and responses as well). There are two main groups of photoreceptors in the human eye: rods and cones. Rods are incapacitated (saturated) during daylight conditions. Cones function as the main tool of visual perception, which is called a photopic response [39].

The radiometric flux parameter (Φ_e, [W]) of light source could be calculated from its photometric parameter (Φ_v, [lm]). The calculations are detailed in several books and articles [36] [38]. The relative spectral sensitivity function ($V(\lambda)$) has to be used as a weighting factor and could be calculated via a non-linear regression (like in [47]). The conversion from luminous flux to radiometric flux data involves a wavelength parameterized component. Because of this, it is necessary to integrate the results of the whole band (the whole emission source spectral band of a component) [36].

Although, this value gives information only about the light source performance. In terms of plant cultivation, one has to be able to control the number of photons per second arriving at an area (plant surface), because photosynthesis is activated (energized) via photons.

The parameter which is used for this purpose is the Photosynthetic Photon Flux Density (PPFD), which is the measure of the number of photons arriving at a unit area during a unit time [$\mu molm^{-2}s^{-1}$].

PPFD could be calculated via the help of spectral irradiance ($E_\lambda(\lambda)$) [37]:

$$PPFD = \frac{10^6}{hcN_A}\int_{400}^{700} E_\lambda(\lambda)d\lambda \tag{1}$$

where h is the Planck's constant (6.626×10^{-34} [Js]), c is the speed of light (3.00×10^8 [ms^{-1}]) and N_A is the Avogadro's number (6.023×10^{23}).

Spectral irradiance ($E_\lambda(\lambda)$ [1]) is a combination of absolute irradiance and spectral power distribution (SPD) [37]:

$$E(\lambda_1,\lambda_2) = \int_{\lambda_1}^{\lambda_2} E_\lambda(\lambda)d\lambda \tag{2}$$

Spectral power distribution is included in the datasheet of the emitters. It is a ratio number over a spectrum band (different wavelengths (λ)).

Absolute irradiance (flux density): E, [Wm^{-2}] could be calculated with the help of radiant intensity, but this solution could be used only if the distance between the emitter and the irradiated surface is large enough. In this way it would not deviate from the inverse-square law (and if we assume that the emitter could be modelled with a Lambertian source model [1]): I, [Wsr^{-1}] [36].

$$E_e = \frac{I_e}{A} = \frac{I_e}{r^2}\ ;\ [W/m^2] = \frac{[W/sr]}{[m^2]} \tag{3}$$

The radiant intensity (I, [Wsr^{-1}]) could be calculated from the radiometric flux (Φ_e, [W]) and with the help of the solid angle (Ω) parameter of the emitter [36].

$$\Omega = \frac{A}{r^2} = \frac{2\pi h}{r} = 2\pi\left[1 - \cos\left(\frac{\alpha}{2}\right)\right] \tag{4}$$

Where α is the viewing angle in terms of emitters (LED). Based on the viewing angle of the emitter, the radiant intensity: I_e, [Wsr^{-1}] for a given waveband could be calculated. The calculation will need the radiometric flux (Φ_e, [W]) and the solid angle (Ω) parameters as well [36]:

$$I_e = \Phi_e / \Omega \,;\; [W/sr] = [W]/[sr] \tag{5}$$

Therefore, we can conclude that under certain conditions (the usage of Lambertian modelling is possible) one can predict the PPFD to a certain extent with an error margin.

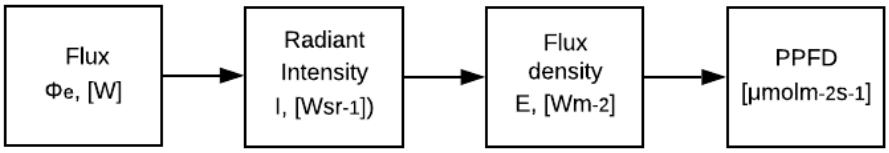

Figure 2
Process of the conversion

It is important to note, that in cases where the Lambertian model is not applicable, one can arrive at false conclusions with the usage of the above model. Safe results could be achieved via the usage of validating measurements on the used emitters.

Usually, PPFD is measured via the usage of (photon) quantum sensors (in [µmolm^{-2}s^{-1}]) such as the LI-COR 190. This machine is quite expensive and gives information only about the PPFD of the location point, where the quantum sensor is placed because it is not feasible and possible to use in a larger plantation area. On the other hand, it is possible to use a much cheaper solution described by Sato and Tsukada: a simple digital camera could give a solution to find the PPFD of the used emitters [37].

4 Plant Specific Light Source Parameters of Analyzed Studies

There are many factors that affect the growth of plants besides light quality. Such parameters are the ground medium, fertilization (amount and mixture type), air exchange ratio, O2 and CO2 levels, air humidity, temperature and light source distance. Several papers – experiments were gathered to find trends among light source settings. The mentioned input variables and the output variables of the examined experiments are not included in this paper due to space limitations. Only the setup parameters of the best results are taken to use them as the base of a statistical analysis to find relationships among them. Mainly non-parametric tests were chosen because of the non-related nature of the examined values. Parametric tests were used in the cases where the data had certain (normal) distribution

Table 3
LED based light source driver setup data of previous projects
Variable abbreviations: F – Frequency, D- Duty cycle, Pho – Photoperiod, I – Light intensity, Pha – Phase shift of different wavelengths, W - Wavelength

Research ID	Type of plant	LED light spectrum	Freq. [Hz]	Duty cycle [%]	PPFD (µmolm^{-2}s^{-1})	Variables	Daily photoperiod
[21] Jao et al. - 2003	Potato - Solanum T. L .	9 red 645 +-20 Nm, 4 blue 460 +-20 Nm	720	50	120	FDPho	16/8
[22] Jao et al. -2004	Potato - Solanum T. L.	9 red 645 +-20 Nm, 4 blue 460 +-20 Nm	720	50	120	FDPho	16/8
[17] Hernández et al.- 2014	Cucumber –C. s. 'Cumlaude'	Red – 661 [Nm]	2500	50**	60	D	18/6
[28] Mori et al. -2002	Lettuce - Lactuca sativa L.	10X20 / 10X40 Clear White	2500	33	50	DPho	24
[25] Kanechi et al. -2016	Lettuce - Lactuca sativa L.	red and blue, 660:455 nm = 4:1	10000	50	200	F	16/8
[33] Pardo et al -2016	Lettuce - Lactuca sativa L.	4 red LEDs (600-650nm)	50	20	240lumens/lamp	W	12/12
[23] Jishi et al. -2012	Lettuce - Lactuca sativa L.	196 white LED	100	75	133,3	FD	16/8
[8] Cho et al. -2013	Lettuce - Lactuca sativa L.	red:blue:white = 5:2:1	2500	25	100	DI	16/8
[16] Harun et al. (2013)	Lettuce - Brassica C.	16 Red 660Nm, 4 Blue 460Nm	not discl.	75	100	Pho	0,45/0,15*24
[48] Timothy et al.., 2016	Lettuce - Lactuca sativa L.	16 Red 660Nm, 4 Blue 460Nm	not disclosed	75	38,8	I	0,45/0,15*24
[19] Iersel et al.2016	Lettuce - Lactuca sativa L.	4X100 W white LED (3000 K)	1000	50 - varied	380 - 520 varied	DI	12/12
[45] Son et al. (2018)	Lettuce - Lactuca sativa L.	red 655 Nm, white 456 Nm + 558 Nm blue 456 Nm LEDs (7:2:1)	1000	75	190	F	12/12
[31] Olvera-Gonzalez et al. -2013	Lettuce - Lactuca sativa L.	30 Blue 460±25 Nm, 170 Red 660±30 Nm, 200 watts/l	0,1	50	750	F	not diclosed
[32] Olvera-Gonzalez et al. -2012	Tomato - Lycopersicon esculentum	30 Blue 460±25 Nm, 170 Red 660±30 Nm, 200 watts/l	1	50	450	F	not diclosed
[46] Tennessen et al.- 1995	Tomato - Lycopersicon esculentum	40 red 668nm,	more than 1000	less than 50	50	FI	16/8
[41] Senol et al. -2016	Carnation – Dianthus sp.	3 Red 615-630 Nm, 3 Red 660 Nm, 3 Blue 455 Nm, 3 Blue 460-475 Nm	5000	50	20	FPha	not diclosed

Table 3
LED based light source driver setup data of previous projects

Research ID	Type of plant	LED light spectrum	Freq. [Hz]	Duty cycle [%]	PPFD (µmolm^{-2}s^{-1})	Notes	Daily photoperiod
[44] Shimada et al. -2011	Thale cress - Arabidopsis thaliana	100 Red 630 Nm, 100 Blue 470Nm, in a unit (Starlet)	2500	45	1800	Pha	12/12
[49] Vejrazka et al. -2012	Microalgae – Chlam. reinhardti	48 Red 630 Nm, 48 Blue 450 Nm	100	50	220	DF	not diclosed
[50] Vejrazka et al. -2013	Microalgae - Chlam. reinhardti	96 Red 630Nm	more than 50	10	115	DF	not diclosed
[20] Jannsen - 2002	Microalgae - Dun. tertiolecta	1452 Red 666 [Nm] 22 LEDs /line	5,31	50	445	F	16/8
[52] Yago et al. - 2012	Micoralgae - Isochrysis g.	White – 400-800 Nm, LED panels	10000	50	104	FI	12/12
[34] Park et al. -2001	Microalgae - Chlorella kessleri	180 Red 680 nm	10-50000	50	78	DF	not diclosed
[42] Sforza et al, - 2012	Microalgae – Nannoch. salina	750nm - LED Light Source SL 3500	30	33	120	FI	not diclosed
[43] Simionato et al. -2013	Microalgae - Nannoch. salina	750nm - LED Light Source SL 3500	30	33	120	ID	not diclosed
[53] Yago et al. - 2014	Micoralgae - Isochrysis g.	470 blue	10000	50	52	W	12/12

The summary of the best results (Table 4), shows that there were several optimal power regulators (mostly PWM) settings for each plant types. The plants are grouped to clusters of sub-types which are noted with addition mark in brackets. The number in the brackets after a plant shows the number of sub-types per plant.

Table 4
Summary of driver settings of different plants about the most efficient settings

Plant type (number of researches)	PWM Duty cycle	Frequency	Light Intensity (PPFD)
Potato (2)	50	720	120
Cucumber (1)	50	2500	120
Lettuce (6+1)	20/25/50	0.1/0.2/50/1000/2500	80/100/125/127/133
Tomato (3)	50	0.1/1/1000	20/100/750
Carnation (1)	50	5000	20
Thale Cress (1)	45	2500	20
Microalgae (2+2+1+1+1)	10/25/50/60	5.3/10/50/100/10000/25000/50000	104-200-220-1800

5 Statistical Analysis

Statistical analysis has been conducted with the help of SPSS (Ver. 20) to find out if there is any significant relationship between the parameters of the experiments. First, the relevant experiments have been collected. Second, the necessary data were collected from the articles. The following variables were used in the analysis: PWM duty cycle, Frequency, Light intensity (PPFD), Light spectrum and Photoperiod. These parameters are suggested by other researches as main parameters as well [6]. As third step, the data were prepared for the descriptive statistical analysis (with the help of section 3 as well). The light spectrum variable is a simplification: it shows the percentage of RED light used in the light sources in the cited experiments. The data were acquired based on Table 3. All these parameters represent the best possible settings to achieve the best results in plant cultivation (under certain conditions). The parameters were "competing" against each other in their own experimental settings. Each of them "won" in their own environment. Therefore, it is not necessary to take into account other variables: all the conditions in the experiment. Fourth, the clustering of the experiments has been done (characterized by their best settings) according to plant types and tested variables. In the end, the statistical analysis has been conducted according to the parametric and non-parametric part of the paper to find significant relationships among variables. Smaller groups (N<6) have been ignored because the reliability of the data they provide is low.

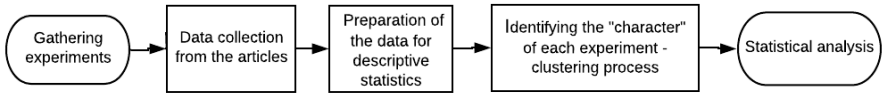

Figure 3
The process of correlation finding across experiments

5.1 Parametric and Non-Parametric Tests

If it is possible to make certain assumptions about a sample, it is possible to apply a parametric test. Such tests are the Pearson's r or ANalysis Of Variance (ANOVA) [54]. Among other assumptions, the population of the test has to have a normal distribution. If it is not the case, the results of the tests will be false. Therefore, the distribution of the data set has to be checked in order to avoid false results. Normal distribution has very specific traits. It is a unimodal, symmetrical curve, with exact areas from the mean after each standard deviation (SD): ±1SD 68.3%, ±2SD 95.4%, ±3SD 99.7% (in SPSS: graphs/leg.dialogs/histogram).

The Pearson's r correlation could be calculated with the following equation [2]:

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[(x_i - \bar{x})^2][(y_i - \bar{y})^2]}} \tag{6}$$

Non-parametric correlational tests had been conducted with Kendall's tau (rank test - correlation signed with r_τ) and also with Spearman's rho (r_ρ) to check for error in the data set. The above mentioned tests were chosen because the variables are continuous and it can be assumed that no specific distribution exists among the specified results of the studies. The statistics take care of the wide variances of the results via giving ranks to the different magnitudes of results. Also, the size of the sample is quite small (N=25 examined studies – 2 omitted experiments because of missing data) – and the distribution of the values are quite wide (therefore Kendall's tau is suggested according to these values as well). Although, Kendall's tau does not take into account tied cases (same values) on dependent or independent variables, therefore most of the times (and in statistical software) Kendall's tau b version is used [2]:

$$r_{\tau_b} = \frac{N_c - N_d}{\sqrt{(N_c + N_d + T_y)(N_c + N_d + T_x)}} \tag{7}$$

Where T_y represents the cases tied on the independent variable while the T_x represents the cases tied on the dependent variable. Spearman' rho could be calculated with the equation [26]

$$r_\rho = 1 - \frac{6\sum_{i=0}^{n} d_i^2}{n(n^2-1)} \tag{8}$$

Where d_i is the difference between the ranks of the corresponding values. The two methods gave similar results, therefore Spearman's rho is more prone to errors and outliers [3]. It is suggested to use the results of Kendall's tau b.

6 Results

There was no significant relationship between the variables of all the experiments (n=25). This was anticipated to happen, because of the differences in plants and the experimental setups. There was no significant relationship among the experiments made with *the tests of Duty cycle* (N=11) and *Light Spectrum* (N=2). Even though significant relationships were found between variables among the experiments made with the *tests of Photoperiod* (N=4), the results were not reliable because of the too low number of cases.

Significant relationship (inverse proportionate) has been found with the Parametric Pearson's r correlation test between light intensity and photoperiod among the variables found in the studies conducted on *lettuce* (N=10; r= -0.824; $p_{\tau b} < 0.003$ (2-tailed)), (Figure 4). Normal distribution was checked via graphical test (Figure 5).

Significant relationship (inverse proportionate) has been found between light intensity and frequency in the studies conducted on *microalgae* (N=8; $r_{\tau b}$=-0.679; (r_ρ= -0.812); $p_{\tau b} < 0.023$ ($p_\rho < 0.014$)). (Figure 6).

Significant relationship (inverse proportionate) has been found between light intensity and frequency in the studies conducted with the *test of frequency* (N=15; $r_{\tau b}$=-0.522; (r_{ρ}= -0.662); $p_{\tau b}$< 0.008 (p_{ρ}< 0.007 (2-tailed)). (Figure 7).

Significant relationship (inverse proportionate) has been found between light intensity and photoperiod in the studies conducted with the *test of Light intensity* (N=7; $r_{\tau b}$=-0.738; (r_{ρ}= -0.833); $p_{\tau b}$< 0.028 (p_{ρ}< 0.02 (2-tailed)). (Figure 8) and also between light intensity and light spectrum (N=7; $r_{\tau b}$=-0.843; (r_{ρ}= -0.935); $p_{\tau b}$< 0.011 (p_{ρ}< 0.001 (2-tailed)). (Figure 9).

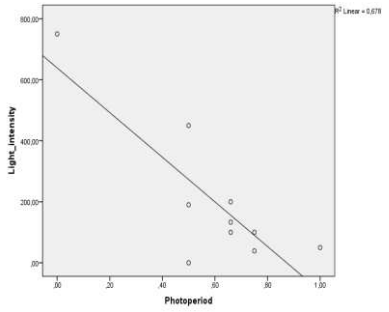

Figure 4
Significant rel. between light intensity & photoperiod on the experiments made on lettuce – Eq. 6.

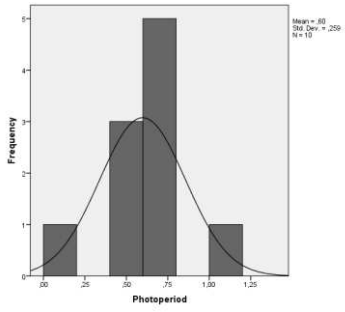

Figure 5
Normal Distribution test of variable Photoperiod (on the experiments made on lettuce

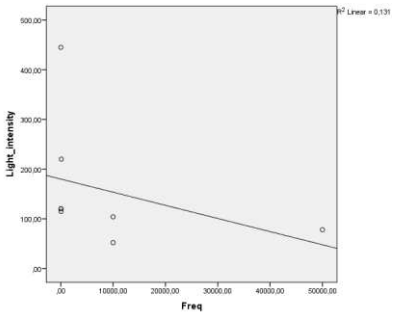

Figure 6
Significant rel. between light intensity and freq. on the experiments made on microalgae - Eq. 7

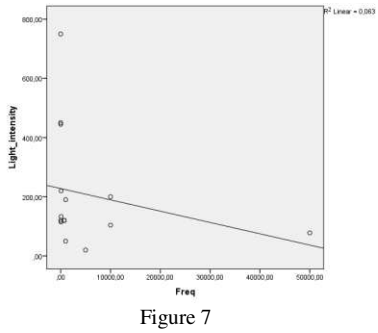

Figure 7
Significant rel. between light intensity and frequency on the experiments made on all the plants conducted with the test of frequency - Eq. 7

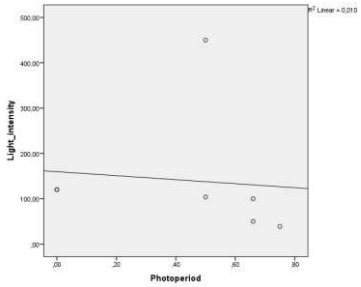

Figure 8
Significant rel. between light intensity and photoperiod on the experiments made on all the plants conducted with the test of light intensity - Eq. 7

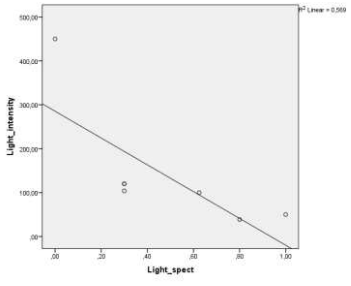

Figure 9
Significant rel. between light intensity and light spectrum on the experiments made on all the plants conducted with the test of light intensity - Eq. 7

Light intensity (PPFD) has significant correlation with other variables almost in every analyzed group. Because it is proportional to the energy used by a light source, the variables correlated with it could be used to reduce the energy costs of a plantation – to enhance efficiency. Photoperiod, frequency and light spectrum are the variables found by this research as the variables affecting the amount of light intensity necessary.

Different plants might require a different wavelength of light (Table 3). With the right spectrum of light, not just plant growth, but energy costs could be lowered as well. Therefore, light equipment could be made with flexible spectral parameters or the light equipment has to be made for a special type of plant. Wavelength and frequency of a light source could be different in the case of different plants (Table 4).

Conclusions

The statistical analysis gave consistent results to other studies [6] and highlighted the correlations for the main parameters of LED light sources.

This type of statistical analysis can support light source design and the setup of agricultural facilities. It has synthesized the results of different experiments. PWM driving of light sources, with optimal settings (frequency, duty cycle, PPFD, spectrum and photoperiod) better yields of plant growth and lower running costs could be obtained.

References

[1] A. V. Arecchi, T. Messadi, and R. J. Koshel, “Field Guide to Illumination”, SPIE, Bellingham, WA, USA., 2007

[2] G. Argyrous, “Statistics for research: With a guide to SPSS” (2nd ed.), London: Sage Publications. 2005

[3] S. Arndt, C. Turvey and N. C. Andreasen, “Correlating and predicting psychiatric symptom ratings: Spearman's r versus Kendall's tau correlation”, Journal of Psychiatric Research, 1999, 33(2), pp. 97-104

[4] I. Ashdown, “Extended parallel pulse code modulation of LEDs”, Proceedings of SPIE, 2006, Vol. 6337, pp. 1-10

[5] N. R. Baker, “Chlorophyll fluorescence: a probe of photosynthesis in vivo”, Annual Review of Plant Biology, 2008, Vol. 59, pp. 89-113

[6] Y. A. Berkovich, I. O. Konovalova, S. O. Smolyanina, A. N. Erokhin, O. V. Avercheva, E. M. Bassarskaya and I. G. Tarakanov, “LED crop illumination inside space greenhouses”, REACH - Reviews in Human Space Exploration, Vol. 6, 2017, pp. 11-24

[7] A. P. Carvalho, S. O. Silva, J. M. Baptista and F. X. Malcata, "Light requirements in microalgal photobioreactors: an overview of biophotonic aspects", Application Microbiology Biotechnol (2011) 89:1275-1288

[8] K. J. Cho, J. Y. Cho, I. S. Park and O. Wook, "Effects of Duty Ratio of Pulsed LED Light on Growth and Photosynthetic Rate of Lettuce Grown in a Plant Factory System", J. Korean Soc. People Plants Environ., 2013, Vol. 16 No. 6: pp.: 427-434

[9] G. Cramer, K. Urano, S. Delrot, M. Pezzotti and K. Shinozaki, "Effects of abiotic stress on plants: a systems biology perspective", BMC Plant Biology 2011, 11, pp. 163

[10] Y. L. Feng, K. F. Cao, and J. L. Zhang, "Photosynthetic Characteristics, Dark Respiration, and Leaf Mass Per Unit Area in Seedlings of Four Tropical Tree Species Grown Under Three Irradiances", Photosynthetica, Vol. 42, 2004, pp. 431-437

[11] K. M. Folta, and S.A. Maruhnich, "Green light: a signal to slow down or stop", Journal of Experimantal Botany, Vol. 58, No. 12, 2007, pp. 3099-111

[12] K. M. Folta, and K. S. Childers, "Light as a Growth Regulator: Controlling Plant Biology with Narrow-bandwidth Solid-state Lighting Systems", HortScience December, Vol. 43, No. 7, 2008, pp. 1957-1964

[13] S. D. Gomkale, and S. T. Zodape, "Effects of pulsed concentrated solar radiation on seeds and plants", Journal of Scientific & Industrial Research, Vol. 58, No. 1, 1999, pp. 11-13

[14] E. Goto, "Effects of light quality on growth of crop plants under artificial lighting", Environment Control in Biology, Vol. 41, ,No. 2, 2003, pp. 121-132

[15] R. N. Hall, G. E. Fenner, J. D. Kingsley, T. J. Soltys, and R. O. Carlson, "Coherent Light Emission From GaAs Junctions", Physical Review Letters, Vol. 9, 1962, pp. 366-369

[16] A. N. Harun, N. N. Ani, R. Ahmad and N. S. Azmi, "Red and blue LED with pulse lighting control treatment for Brassica chinensis in Indoor farming", Presented at 2013 IEEE Conference on Open Systems (ICOS), Dec.2-4, Sarawak, Malaysia

[17] R. Hernández, A. Dragotakes and C. Kubota, "Pulsing Effects of Supplemental LED Lighting on Cucumber Seedlings Growth and Morphology in Greenhouse", Acta Hort. (ISHS) 2014, 1037: p875-880

[18] N. Holonyak, and S. F. Bevacqua, "Coherent (visible) Light Emission from Ga(As1-xPx) Junctions", Applied Physical Letters, Vol. 1, No. 82, 1962

[19] M.W. Iersel, E. Mattos, G. Weaver, R.S. Ferrarezi, M.T. Martin and M. Haidekker, "Using chlorophyll fluorescence to control lighting in controlled environment agriculture". Acta Hort. (ISHS), 2016, Vol. 1134, pp. 427-434

[20] M. G. J. Janssen, "Culivation of microalgae: effect of light/dark cycles on biomass yield", Doctoral Thesis, Wageningen University, Wageningen, The Netherlands, 2002

[21] R.-C. Jao and W. Fang, "An Adjustable Light Source for Photo-Phyto Related Research and Young Plant Production", Applied Engineering in Agriculture, 2003, Vol. 19(5): pp.: 601-608

[22] R.-C. Jao and W. Fang, "Effect of frequency and duty ratio on the growth of potato plantlets in vitro using light emitting diodes", HortScience, 2004, Vol. 39(2): pp.: 375-379

[23] T. Jishi, K. Fujiwara, K. Nishino and A. Yano, "Pulsed Light at Lower Duty Ratios with Lower Frequencies is Less Advantageous than Continuous Light for CO2 Uptake in Cos Lettuce", Journal of Light & Visual Environment, 2012, Vol. 36, No. 3, pp. 88-93

[24] Cs. Kárász, J. Kopják, "Comparative study about LED driving methods and feedback system for indoor plant cultivation", 17th IEEE International Symposium on Computation Intelligence and Informatics, Transactions on Power Electronics, Budapest, Hungary, 2016, pp. 219-224

[25] M. Kanechi, "Growth and Photosynthesis under pulsed lighting In Photosynthesis - From Its Evolution to Future Improvements in Photosynthetic Efficiency Using Nanomaterials", Edited by Juan Cristóbal García Cañedo and Gema Lorena López-Lizárraga, IntechOpen, 2018

[26] M. G. Kendall, "Rank correlation methods", (1970), London, Griffin

[27] J. Masojidek, M. Koblizek and G. Torzillo, "Photosynthesis in microalgae" In: Richmond A (ed) Handbook of microalgal culture: biotechnology and applied phycology. Blackwell Science, Oxford, 2004, pp. 20-39

[28] Y. Mori, M. Takatsuji and T. Yashuoka, "Effects of Pulsed White LED Light on the Growth of Lettuce", Shokubutsu Kojo Gakkaishi, 2002, 14(3): pp.:136-140

[29] A. Möglich, X. Yang, R. A. Ayers and K. Moffat, "Structure and function of plant photoreceptors", Annual Review of Plant Biology, Vol. 61, 2010, pp. 21-47

[30] L. Nedbal, V. Tichy, F. Xiong and J.U. Grobbelaar, "Microscopic green algae and cyanobacteria in high-frequency intermittent light"., Journal of Applied Physics, Vol. 8, 1996, pp. 325-333

[31] E. Olvera-Gonzalez, D. Alaniz-Lumbreras, R. Ivanov-Tsonchev, J. Villa-Hernández, C. Olvera-Olvera, E. González-Ramírez, M. Araiza-Esquivel, V. Torres-Argüelles and V. Castaño, "Intelligent lighting system for plant growth and development", Computers and Electronics in Agriculture, Vol. 92, 2012, pp. 48-53

[32] E. Olvera-Gonzalez, D. Alaniz-Lumbreras, R. Ivanov-Tsonchev, J. Villa-Hernández, C. Olvera-Olvera, E. González-Ramírez, M. Araiza-Esquivel, V. Torres-Argüelles and V. Castaño, "Chlorophyll fluorescence emission of tomato plants as a response to pulsed light based LEDs", Plant Growth Regulation, 2013, Volume 69, Issue 2, pp 117-123

[33] G. P. Pardo, S. T. Velázquez, A. Cruz-Orea, C. H. Aguilar, F. Arturo, D. Pacheo and F. R. Martínez, "Pulsed led light in germination and growth of lettuce seeds", Bothalia Journal, 2017, Vol. 46, No. 4, Apr 2016, pp. 13-26

[34] K.-H. Park, and C. G. Lee, "Effectiveness of flashing light for increasing photosynthetic efficiency of microalgal cultures over a critical cell density", Biotechnology and Bioprocess Engineering, 2001, Vol. 6(3): pp. 189-193

[35] A. C. Parr, "The Candela and Photometric and Radiometric Measurements", Journal of Research of the National Institute of Standards and Technology, Vol. 106, 2000, pp. 151-186

[36] A. Ryer, "Light Measurement Handbook", Int. Light, Newburyport, MA, 2000

[37] A. Sato and M. Tsukada, "Measuring method for photosynthetic photon flux density using a digital camera", Acta Hortic. Vol. 1170, No. 125, 2017, pp. 973-980

[38] E. F. Schubert, "Light-emitting diodes", 2nd Ed., Cambridge Univ. Press, NY, 2006

[39] J. Schwiegerling, "Field Guide to Visual and Ophthalmic Optics", SPIE Press, Bellingham, Washington, USA, 2004

[40] R. Senol and K. Taşdelen, "A New Approach for LED Plant Growth Units", Acta Polytechnica Hungarica, 2014, Vol. 11, No. 6

[41] R. Senol, S. Kilic and K. Tasdelen, "Pulse timing control for LED plant growth unit and effects on carnation", Computers and Electronics in Agriculture, 2016, Vol. 123, pp.: 125-134

[42] E. Sforza, D. Simionato, G. M. Giacometti, A. Bertucco and T. Morosinotto, "Adjusted Light and Dark Cycles Can Optimize Photosynthetic Efficiency in Algae Growing in Photobioreactors", . PLoS ONE, 2012, Vol. 7(6), p1-10

[43] D. Simionato, S. Basso, G. M. Giacometti and T. Morosinotto, "Optimization of light use efficiency for biofuel production in algae", Biophysical Chemistry, Vol. 182, No. 1, 2013, pp. 71-78

[44] A. Shimada and Y. Taniguchi, "Red and blue pulse timing control for pulse width modulation light dimming of light emitting diodes for plant cultivation", Journal of Photochemistry and Photobiology B: Biology, 2011, Vol. 104, No, 3, pp.: 399-404

[45] K.-H. Son, S.-R. Lee and M.-M. Oh, "Comparison of lettuce growth under continuous and pulsed irradiation using light-emitting diodes", The Korean Society of Horticulture, Horticultural Science and Technology, 2018, Vol. 36, No. 4 pp. 542-551

[46] D. J. Tennessen, R. J. Bula and D. Sharkey, "Efficiency of photosynthesis in continuous and pulsed light emitting diode irradiation", Photosynthesis Research, 1995, Volume 44, Issue 3, pp. 261-269

[47] Thorlabs, "Light emitting diode technologies", Thorlabs Inc, .Newton, New Jersey, United States p. 506

[48] M. Timothy, C. Ira and W. John, "Horticulture of lettuce (Lactuvasativa L.) using red and blue led with pulse lighting treatment and temperature control in SNAP hydroponics setup", Journal Teknologi (Sciences & Engineering), 2016, Vol. 78, No. 9, pp. 67-71

[49] C. Vejrazka, M. Janssen, M. Streefland and R. H. Wijffels, "Photosynthetic efficiency of Chlamydomonas reinhardtii in attenuated, flashing light", Biotechnology and Bioengineering; 2012, Vol. 109, No. 10, pp.: 2567-74

[50] C. Vejrazka, M. Janssen, G. Benvenuti, M. Streefland and R. H. Wijffels, "Photosynthetic efficiency and oxygen evolution of Chlamydomonas reinhardtii under continuous and flashing light", Applied Microbiology&Biotechnology, 2013, Vol. 97, pp: 1523-1532

[51] J. Whitmarsh and Govindjee, "Photosynthesis". In: Encyclopedia of Applied Physics. Wiley; 1995, Vol. 13, pp. 513-532

[52] T. Yago, H. Arakawa, K. Fukui, B. Okubo, K. Akima, S. Takeichi, Y. Okumura and T. Morinaga, "Effects of flashing light from light emitting diodes (LEDs) on growth of the microalga Isochrysis galbana", African Journal of Microbiology Research 2012, Vol. 6(30), pp. 5896-5899

[53] T. Yago, H. Arakawa, K. Akima, Y. Okumura and T. Morinaga, "Effects of flashing light-emitting diode (LED) of several colors on the growth of the microalga Isochrysis galbana", African journal of microbiology research., 2014, Vol. 8., pp. 3815-3820

[54] J. Zheng, W. Zhuang, N. Yan, G. Kou, H. Peng, C. McNally, D. Erichsen, A. Cheloha, S. Herek and C. Shi, "Classification of HIV-1-mediated neuronal dendritic and synaptic damage using multiple criteria linear programming", Neuroinformatics. 2004, Vol. 2(3), pp. 303-26.

Analysis of Heat Release Dynamics in a Compression-Ignition Engine with Multiple Fuel Injection

Zdzisław Chłopek, Jakub Lasocki

Warsaw University of Technology, Faculty of Automotive and Construction Machinery Engineering, Institute of Vehicles, Department of Combustion Engines, Narbutta 84, Warsaw, 02-524, Poland, e-mails: zdzislaw.chlopek@pw.edu.pl, jakub.lasocki@pw.edu.pl

Abstract: The heat release characteristics of the combustion process in a four-stroke compression-ignition engines depend strongly on the fuel injection strategy. In the case of multiple fuel injection, deviations from the classic heat release model applicable to a single injection are observed. This paper presents the results of empirical tests and their analysis for a single-cylinder compression-ignition research engine with double injection of fuel. The influence of fuel injection pattern on the indicated pressure has been analyzed, in particular the derivative of indicated pressure in respect of the crank angle. The processes of fuel supply and heat release have been investigated in time domain. A correlation analysis of voltage controlling the opening of injector and unit heat release rate have been presented. The delay angle and delay time of unit heat-release rate in respect of the voltage controlling the opening of the injector have been determined. The heat release for each injected fuel dose has been analyzed.

Keywords: heat release; compression-ignition engine; multiple fuel injection; ignition delay

1 Introduction

It is widely known that fuel injection strategy plays a dominant role in determining the performance of combustion engine, especially in the case of direct-injection fuel supply systems of four-stroke compression-ignition engines [14, 15]. Multiple or split fuel injection (i.e. the division of the fuel dose into several parts injected during one working cycle) has been applied in automotive engines for some time now. It allows a more precise control of the heat release characteristics of the combustion process and positively influences thermal efficiency of the engine as well as pollutant emission [3, 9, 10, 12, 23].

The course of the combustion process in compression-ignition engines is strongly influenced by phenomena associated with fuel injection and the formation of combustible mixture [24]. These phenomena determine the autoignition delay and the course of heat release [2, 22, 25]. They have been sufficiently investigated in the case of single fuel injection, providing analytical description that allows to generalize the formulated conclusions [13, 17]. However, in the case of multiple fuel injection, the combustion of the second and subsequent doses of fuel proceeds differently than the first dose, for which spontaneous combustion must occur. For subsequent injections there are significant differences in the phenomena associated with ignition [16]. Firstly, mixture formation takes place under conditions other than during the first injection: at a higher temperature and higher pressure of the medium, and in an environment comprising a mixture of air, fuel vapor and exhaust components, wherein mixture formation is accompanied by already spreading flames. Secondly, the initiation of ignition occurs not only due to pre-ignition reactions, but also to flames that have already started to spread, originating from the spontaneous combustion of the first dose, and, in the case of subsequent doses, from the burning of the previous dose. There is a need, therefore, to study the dynamics of ignition delay to the subsequent fuel doses. Although there are many studies in this field, both empirical [2, 10, 22, 23, 25] and theoretical [11, 17-19, 22, 23, 25], none has so far presented generalized knowledge of the dynamics of fuel combustion processes for multiple injection.

This paper presents detailed analysis of heat release dynamics in a compression-ignition engine with double fuel injection. The study was based on the results of empirical tests performed with the use of a single-cylinder compression-ignition research engine. The objective was to provide the detailed insight into the effects of splitting the fuel dose on heat release, its magnitude and rate of change.

2 Materials and Methods

Empirical tests that served as the basis for the considerations contained in this work have been described in the previous paper [6]. It presents the detailed technical specifications of the engine, the test stand and instrumentation, as well as characteristics of the fuel and procedures of the test program. Therefore, only basic information is given in this section.

Tests were carried out on the AVL Single Cylinder Test Bed [4] consisting of a single-cylinder compression-ignition research engine AVL 5402 (Table 1), a set of exhaust-gas analyzers, and instrumentation controlling the operation of the entire system. Data has been collected and processed using the software AVL PUMA and AVL CONCERTO.

The test program included engine operating in static states, at the external speed characteristic with engine rotational speed within the range of 1200-3600 rpm in steps of 400 rpm. The basic measurements involved the following quantities: rotational speed – n, torque – M_e, indicated pressure – p_g, recorded in crank angle domain – α, voltage of the injector opening control signal – U as well as other quantities, which were not included in this work.

Table 1
Technical specifications of the AVL 5402 engine [6]

Specification	Data
Number of cylinders	1
Bore	85.01 mm
Stroke	90.00 mm
Displacement	511.00 cm^3
Combustion type	Compression ignition
Valve system	4 valves
Compression ratio	17.0 ÷ 17.5
Fueling system	Direct injection, Common Rail
Maximum effective power	6 kW
Rated engine speed	4200 min^{-1}
Injection pressure	180 MPa

The data on pressure inside engine cylinder was obtained by recording of 20 traces of indicated pressure vs. the crank angle at each engine working point. The measurements were performed with a resolution of 1 deg. of crank rotation, except for the crank angle in the range of -30–90 deg., corresponding to active combustion near the Top Dead Centre (TDC), where a resolution equaled 0.1 deg.

In order to reduce the contamination of recorded signals with high-frequency noise, processing of the signals was performed, consisting in synchronous averaging [5, 20]. The rough estimator of numerical differentiation has been low-pass filtered to reduce the share of high-frequency noise in the signal [7, 8]. For filtering, a non-recurring filter was applied five times:

$$\overline{y}_i = \frac{1}{11}\sum_{j=i-5}^{i+5} y_j \tag{1}$$

where y – rough estimator, $\overline{y}$ – filtered estimator.

The engine was supplied with diesel oil. Its properties are described in [6].

3 Results and Discussion

3.1 Analysis of the Course of Indicated Pressure

Figure 1 shows the influence of double fuel injection on the indicated pressure and the derivative of the indicated pressure at the engine working point corresponding to the maximum torque (1600 rpm). In contrast, Figure 2 shows the typical curves of the indicated pressure and the derivative of the indicated pressure for the single fuel injection.

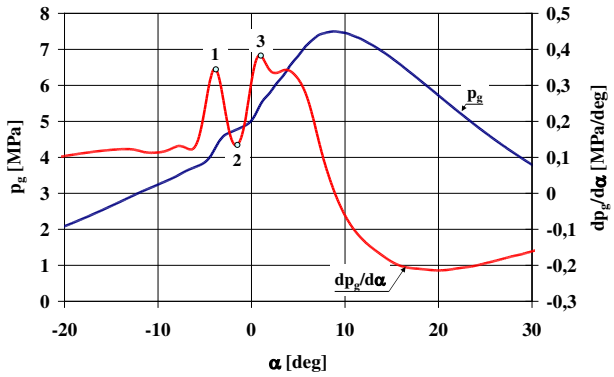

Figure 1
The indicated pressure and the derivative of the indicated pressure at the engine working point corresponding to maximum torque (1600 rpm)

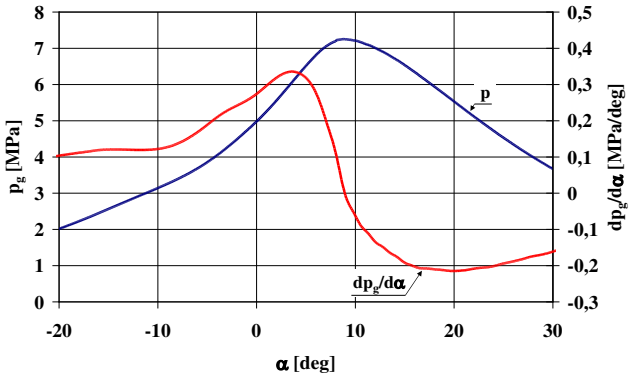

Figure 2
Typical curves of the indicated pressure and the derivative of the indicated pressure for engine with single fuel injection

Figure 3 illustrates the relationship between the values of three extrema of the derivatives of the indicated pressure and the engine rotational speed. The average value and standard deviation of these extrema for each rotational speed are also plotted in the graph.

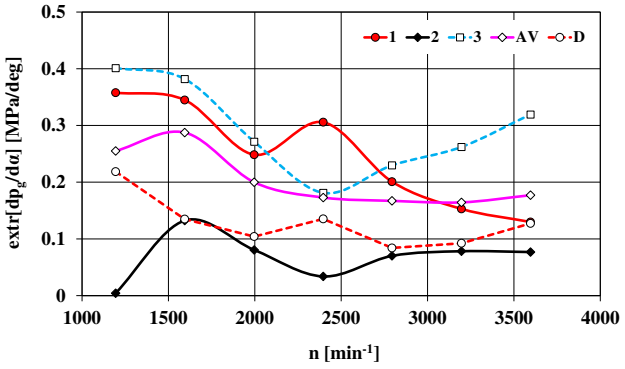

Figure 3
The values of three extrema of the derivative of the indicated pressure and their average value – AV and standard deviation – D, depending on the engine rotational speed

3.2 Analysis of the Processes of Fuel Supply and Heat Release

On the basis of the indicated pressure and the information associated with the parameters of engine and fuel, the following quantities have been determined: the unit heat (understood as the heat related to the volume, i.e. the cylinder displacement), the unit heat-release rate, and the temperature of medium. For this purpose, AVL CONCERTO and AVL BOOST algorithms have been applied. Figure 4 presents the courses of those quantities along with the indicated pressure and the voltage of the injector opening control signal.

Figure 5 shows diagrams of crank angles corresponding to the important phenomena of fuel supply and heat release in the cylinder. Diagram (a) depicts the case where for the second fuel dose there is one maximum of the rate of unit heat release while diagram (b) – the case where the number of maxima is larger (three in analyzed tests). The crank angles shown in Figure 5 correspond to (notation (a) or (b) indicates that the crank angle is specific exclusively to diagram (a) or (b)): the autoignition delay – DIGN, the start of the injection control signal – US1, the maximum value of the control signal in the first injection – UMAX1, the end of the control signal in the first injection – UE1, the length of the control signal in the first injection – U1, the start of the control signal in the second injection – US2, the maximum value of the control signal in the second injection – UMAX2 (a), first maximum value of the control signal of the second injection – UMAX21 (b),

second maximum value of the control signal of the second injection – UMAX22 (b), third maximum value of the control signal in the second injection – UMAX23 (b), average value of the maximum values of the control signal in the second injection – UMAXAV (b), the end of the control signal in the second injection – UE2, the length of the control signal in the second injection – U2, the start of the heat release in the first injection – QS1, maximum value of unit heat-release rate in the first injection – QMAX1, the start of the heat release in the second injection – QS2, the maximum value of the unit heat-release rate in the second injection – QMAX2 (a), first maximum value of the unit heat-release rate in the second injection – QMAX21 (b), second maximum value of unit heat-release rate in the second injection – QMAX22 (b), third maximum value of the unit heat-release rate in the second injection – QMAX23 (b), the average value of the maximum values of the unit heat-release rate in the second injection – QMAXAV (b), the delay to the ignition of first dose – DS1, the delay to the maximum value of the unit heat-release rate in relation to the maximum value of the control signal in the first injection – DMAX1, the delay to the ignition of the second dose – DS2, the delay to the maximum value of the unit heat-release rate in relation to the maximum value of the control signal in the second injection – DMAX2 (a), the delay to the first maximum value of the unit heat-release rate in relation to the maximum value of the control signal in the first injection – DMAX21 (b), the delay to the second maximum value of the unit heat-release rate in relation to the maximum value of the control signal in the second injection – DMAX22 (b), the delay to the third maximum value of the unit heat-release rate in relation to the maximum value of the control signal in the second injection – DMAX23 (b), the average value of the maximum values of unit heat-release rate in relation to the maximum value of the control signal in the second injection – DMAX2AV (b).

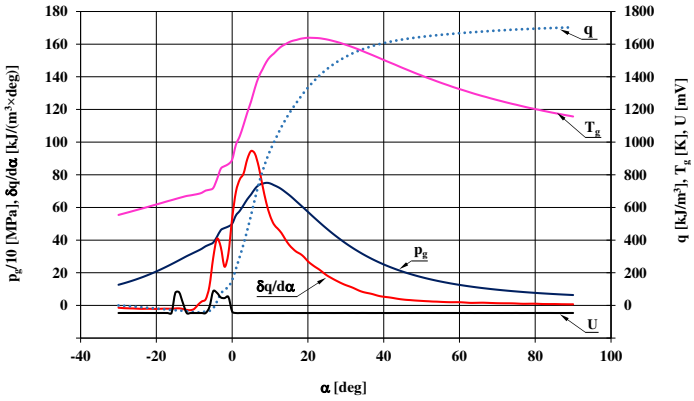

Figure 4

Indicated pressure – p_g, temperature of medium – T_g, unit heat-release rate – $\delta q/d\alpha$, unit heat – q and the voltage of the injector opening control signal – U at the engine working point corresponding to maximum torque (1600 rpm)

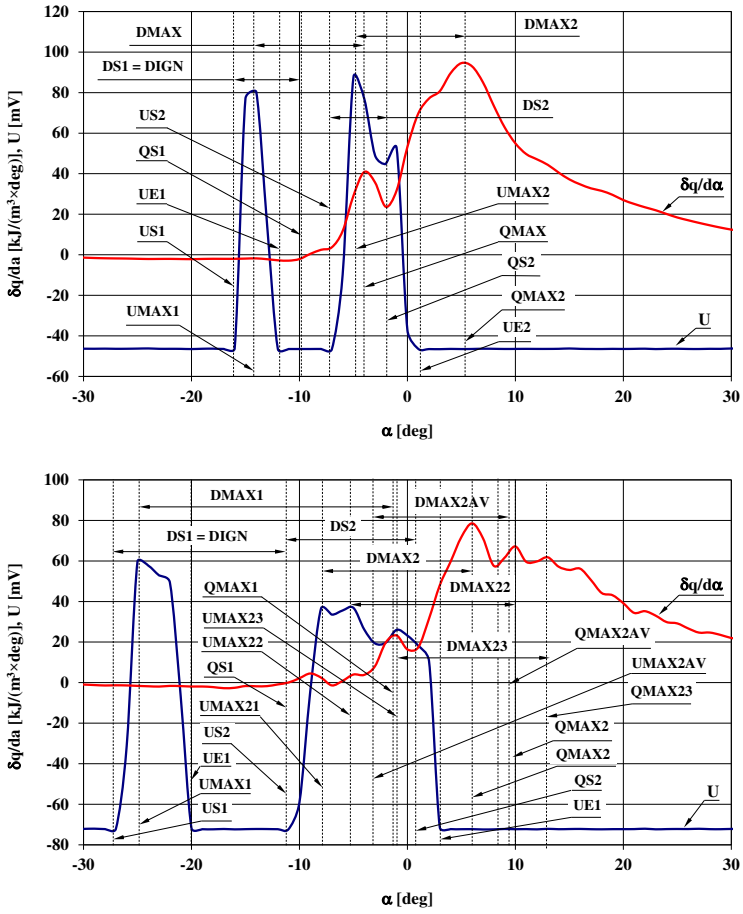

Figure 5

The rate of unit heat release – δq/dα and the voltage of the injector opening control signal – U at the engine working point corresponding to maximum torque (1600 rpm) for one maximum (a) and three maxima (b) of the rate of unit heat release (see the text for the description of the angles)

Figure 6 shows the angles corresponding to the start of injection, autoignition and autoignition delay as a dependence on the engine rotational speed. The start of the injection angle is an increasing function of the rotational speed. This is due to the fact that it is necessary to initiate combustion in the appropriate phase of the engine cycle, and for increasing the rotational speed, the shorter time corresponds to the same angle of rotation. The start of the combustion angle is practically constant in the rotational speed domain, while the autoignition-delay angle is an obviously decreasing function – approximately linearly.

Figure 7 shows the angle of the start of the injection control signal, the angle of the maximum value of the control signal in the first injection, the angle of the end of the control signal in the first injection, and the angle of the length of the control signal in the first injection as a dependence on the engine rotational speed. The angle of the length of the control signal in the first injection is practically constant in the rotational-speed domain, while other angles decrease, which corresponds approximately to the constancy of the processes in question in the timescale.

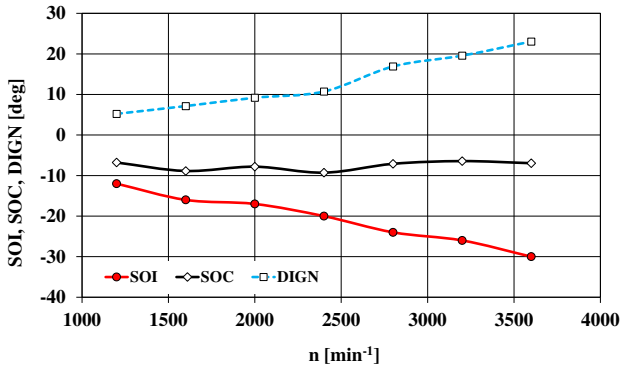

Figure 6

The start of the injection angle – SOI, the autoignition angle – SOC and the autoignition-delay angle – DIGN as a dependence on the engine rotational speed

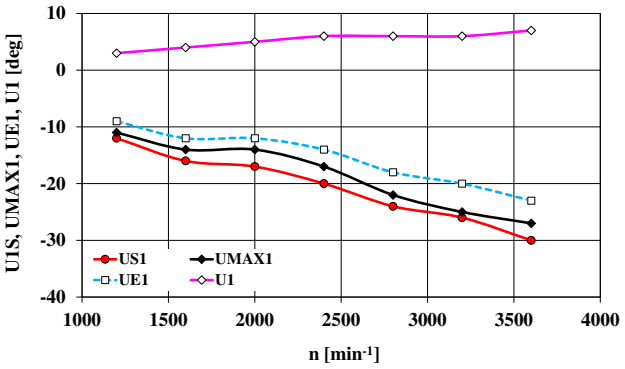

Figure 7

The angles of: the start of the injection control signal – US1, the maximum value of the control signal in the first injection – UMAX1, the end of the control signal in the first injection – UE1 and the length of the control signal in the first injection – U1, as a dependence on the engine rotational speed

Figure 8 is an illustration of the dependence of the angles corresponding to the phenomena associated with the second injection on the engine rotational speed. It

can be seen that the angle of the length of the control signal in the second injection increases together with the increase in the rotational speed, while injection end of the second stage is practically constant. Other angles are weakly dependent on the rotational speed with a tendency to decreasing dependence.

Figure 9 shows the relationship between the angles characterizing heat release inside engine cylinder and engine rotational speed. Analysis of the plotted data allows to evaluate this relationship as weak – no clear trend can be identified.

Figure 10 shows the relationship between the length of the control signal of subsequent injections and engine rotational speed. These are increasing functions, which corresponds to the increase of fuel dose with increasing rotational speed.

Figures 11 and 12 show the dependence of the angles of ignition delay and the maximum value of the unit heat-release rate delay in relation to the maximum value of the control signal in each phase of ignition on the engine rotational speed. These are increasing functions of the rotational speed.

Figures 13 and 14 show the relationship between the angle of the maximum value of the unit heat-release rate and the angle of the maximum value of the control signal in each phase of ignition. These are quantities weakly correlated to each other. In the first case, the Pearson linear-correlation coefficient [21] is -0.494, and the probability that the hypothesis of the absence of a correlation would not be rejected is 0.260, while in the second case –0.035 and 0.941, respectively. These results can be interpreted as a lack of linear correlation between the analyzed sets.

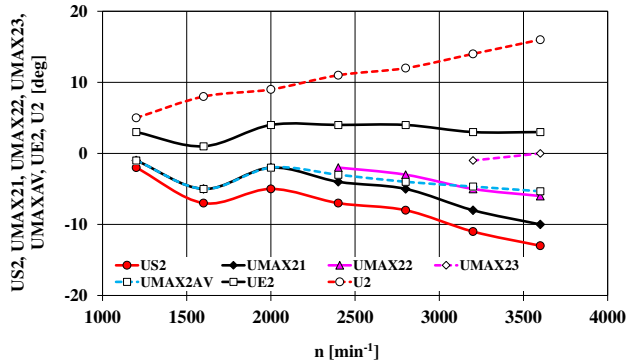

Figure 8

The angles of: the start of the control signal in the second injection – US2, the first maximum value of the control signal in the second injection – UMAX21, the second maximum value of the control signal in the second injection – UMAX22, the third maximum value of the control signal in the second injection – UMAX23, the average value of the maximum values of the control signal in the second injection – UMAXAV, the end of the control signal in the second injection – UE2 and the length of the control signal in the second injection – U2 as a dependence on the engine rotational speed

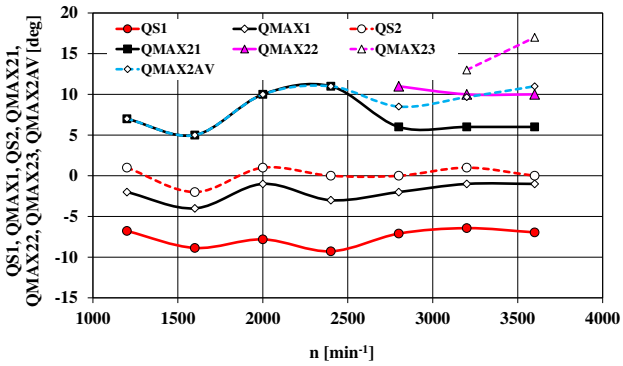

Figure 9

The angles of: the start of heat release in the first injection – QS1, the maximum value of unit heat-release rate in the first injection – QMAX1, the start of heat release in the second injection – QS2, the first maximum value of unit heat-release rate in the second injection – QMAX21, the second maximum value of unit heat-release rate in the second injection – QMAX22, the third maximum value of unit heat-release rate in the second injection – QMAX23 and the average value of the maximum values of unit heat-release rate in the second injection – QMAXAV as a dependence on the engine rotational speed

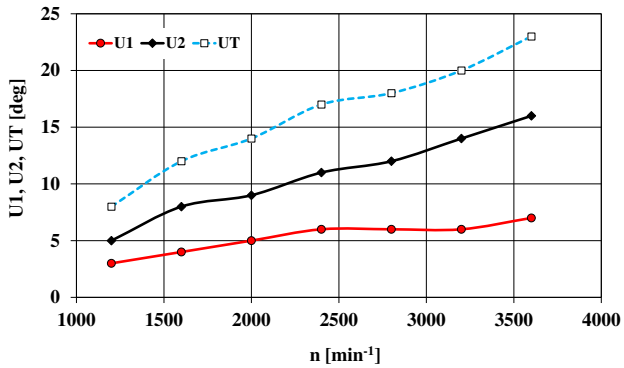

Figure 10

The angles of: the length of the control signal in the first injection – U1, the length of the control signal in the second injection – U2 and the total length of the control signal of injection – UT as a dependence on the engine rotational speed

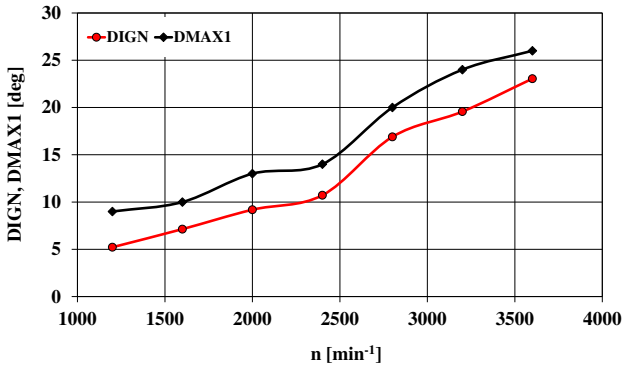

Figure 11

The angles of: the autoignition delay – DIGN and the delay to the maximum value of the unit heat-release rate in relation to the maximum value of the control signal in the first injection – DMAX1 as a dependence on the engine rotational speed

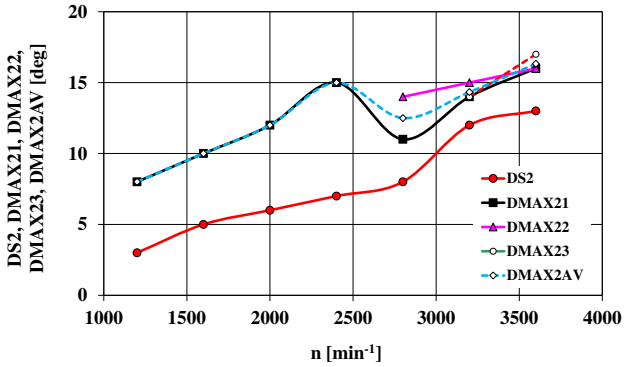

Figure 12

The angles of: the delay to the ignition of the second dose – DS2, the delay to the first maximum value of the unit heat-release rate in relation to the maximum value of the control signal in the first injection – DMAX21, the delay to the second maximum value of unit heat-release rate in relation to the maximum value of the control signal in the second injection – DMAX22, the delay to the third maximum value of the unit heat-release rate in relation to the maximum value of the control signal in the second injection – DMAX23, the average value of the maximum values of the unit heat-release rate in relation to the maximum value of the control signal in the second injection – DMAX2AV as a dependence on the engine rotational speed

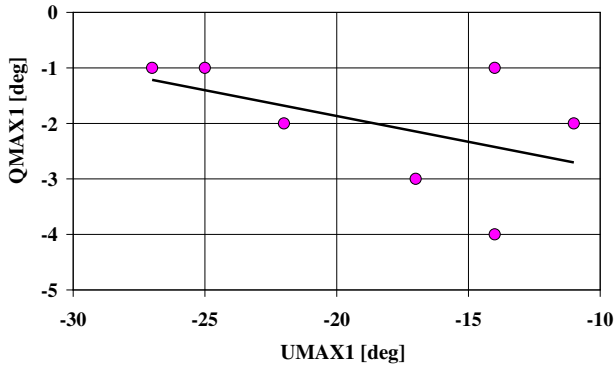

Figure 13
The relationship between the angle of the maximum value of the unit heat-release rate in the first injection – QMAX1 and the angle of the maximum value of the control signal in the first injection – UMAX1

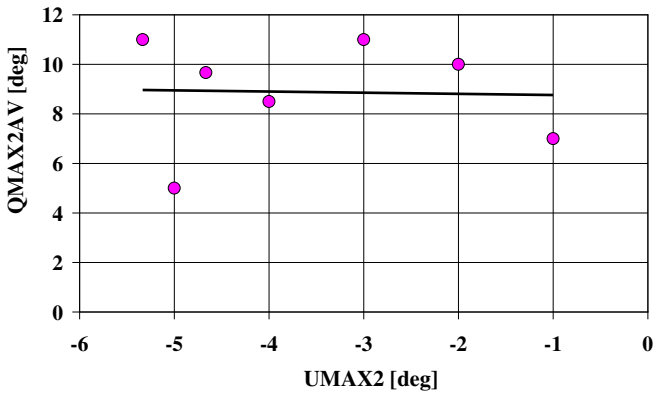

Figure 14
The relationship between the angle of the average value of maximum values of the unit heat-release rate in the second injection – QMAXAV and the angle of the average maximum values of the control signal in the second injection – UMAXAV

Since the rate of heat release is determined by the flow of fuel supplied to the cylinder, a correlation analysis of the processes of injector opening control voltage and unit heat-release rate has been performed. The algorithm using a fast Fourier transformation implemented in the Statistica 6.1 software has been used to determine the cross-correlation function of processes [5, 20]. To investigate it, an assumption about the stationarity of these processes was adopted [5, 20]. In this algorithm, the Fourier transform method is calculated using the mutual spectral power density of the studied processes and then the function of the cross

correlation of these processes is determined by the method of the inverse fast Fourier transform.

Figures 15 and 16 show the cross-correlation of standardized voltage signals controlling the opening of the injector and the rate of the unit heat release in the domain of the crank angle offset for maximum torque and power output.

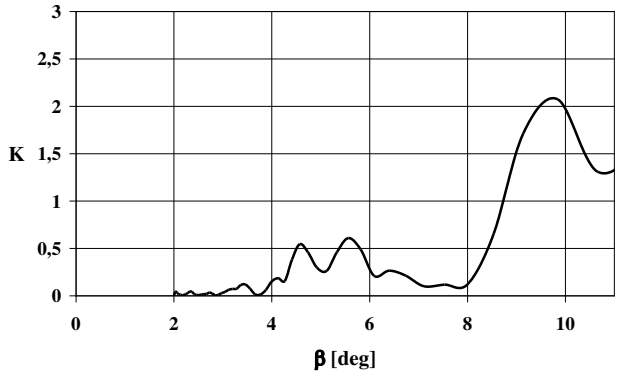

Figure 15
Cross correlation – K of standardized voltage signals controlling the opening of the injector – U and unit heat-release rate – $\delta q/d\alpha$ in the domain of the crank-angle offset – β for maximum torque (n = 1600 rpm)

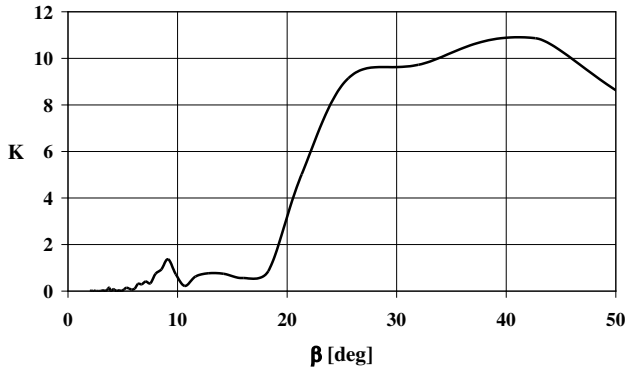

Figure 16
Cross correlation – K of the standardized voltage signals controlling the opening of the injector – U and unit heat-release rate – $\delta q/d\alpha$ in the domain of the crank-angle offset – β for maximum power output (n = 1600 rpm)

The first distinct maxima correspond to the estimated delay to the signals. Figure 17 shows the speed characteristics of the delay angle and Figure 18 – of the delay

time, for the signal of the unit heat-release rate in respect of the signal of the control voltage of the injector opening. The delay angle of the unit heat-release rate in respect of the signal of the control voltage of the injector opening is an increasing function of the rotational speed. This angle is in the range of 10-40 deg. The delay time of the unit heat-release rate in respect of the signal of the control voltage of the injector opening is weakly dependent on the rotational speed – this time depends mainly on the kinetics of the pre-ignition reactions [13, 22].

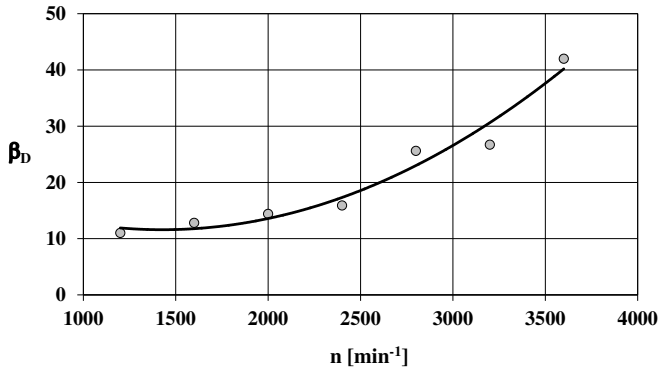

Figure 17
The speed characteristics of delay angle – β_D of the signal of the unit heat-release rate in respect of the signal of the control voltage of injector opening

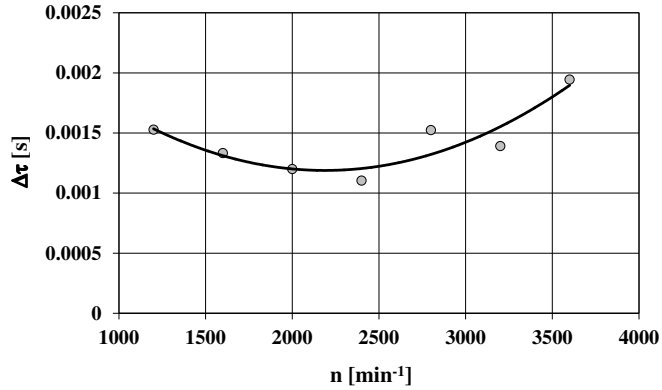

Figure 18
The speed characteristics of delay time – $\Delta\tau$ of the signal of the unit heat-release rate in respect of the signal of control voltage of injector opening

3.3 Analysis of the Heat Release

The analysis of the heat release in the cylinder involved the unit heat-release rate, along with the unit heat for the two doses of the fuel injected and in various parts of the second injection, restricted by the local minima of the unit heat-release rate. Figure 19 shows schematic indications for the analysis of the heat release.

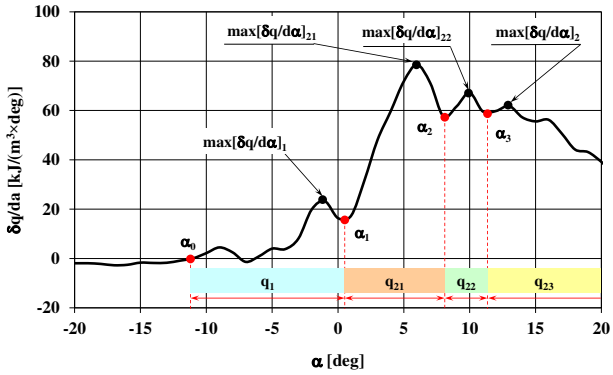

Figure 19

The unit heat-release rate – the identification of the maximum values of the unit heat-release rate and the angles limiting parts of the ith-phase of injection

The unit heat in two stages of the fuel injection and in different parts of the second injection, as well as the total unit heat – q_T – were calculated according to the formulas

$$q_1 = \int_{\alpha_0}^{\alpha_1} \frac{\delta q}{d\alpha} d\alpha; \quad q_{21} = \int_{\alpha_1}^{\alpha_2} \frac{\delta q}{d\alpha} d\alpha; \quad q_{22} = \int_{\alpha_2}^{\alpha_3} \frac{\delta q}{d\alpha} d\alpha; \quad q_{23} = \int_{\alpha_3}^{\alpha_E} \frac{\delta q}{d\alpha} d\alpha \tag{2}$$

$$q_T = q_1 + q_{21} + q_{22} + q_{23}$$

where α_E – the angle of the conventional end of combustion.

Figure 20 shows the maximum value of the unit heat-release rate in two stages of fuel injection and in different parts of the second injection as the dependence on the engine rotational speed. The maximum unit heat-release rate in the second injection is much higher, by more than twice, than in the first injection. The maximum value of the unit heat-release rate in the first injection has a decreasing tendency in the domain of rotational speed. For the rotational speed, for which there are no more maxima of unit heat-release rates in the second injection, the maximum value of the unit heat-release rate is also a decreasing function of the rotational speed. For the higher rotational speeds the maximum value of the unit heat-release rate in the first part of the second injection increases together with the

increase in the rotational speed, while in the subsequent parts is a decreasing function.

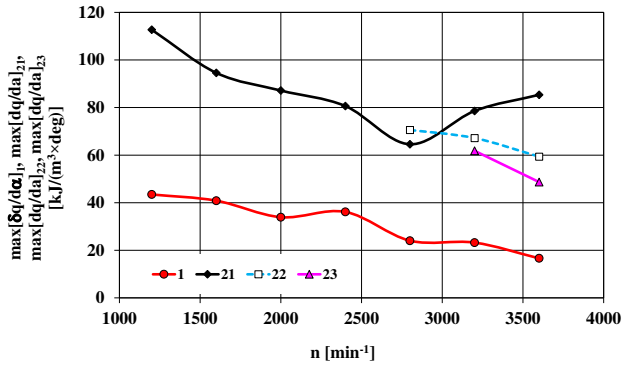

Figure 20

The maximum value of the unit heat-release rate – max[δq/dα] (1 – in the first injection, 21 – in the first part of the second injection, 22 – in the second part of the second injection and 23 – in the third part of the second injection) as a dependence on the engine rotational speed

Figure 21 shows the unit heat, while Figure 22 the relative unit heat, related to the total unit heat in the two stages of fuel injection and in different parts in the second injection, as a dependence on the engine rotational speed.

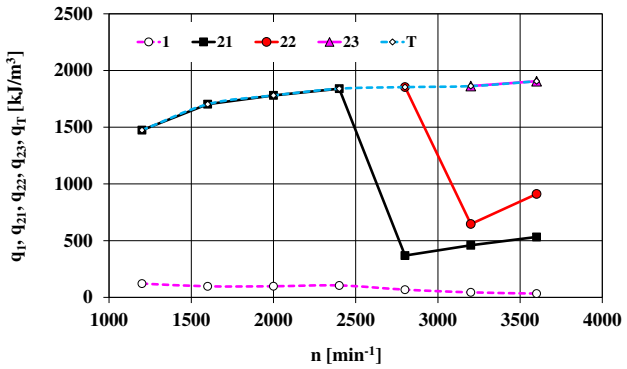

Figure 21

Unit heat – q (1 – in the first injection, 21 – in the first part of the second injection, 22 – in the second part of the second injection, 23 – in the third part of the second injection and T – the total unit heat) as a dependence on the engine rotational speed

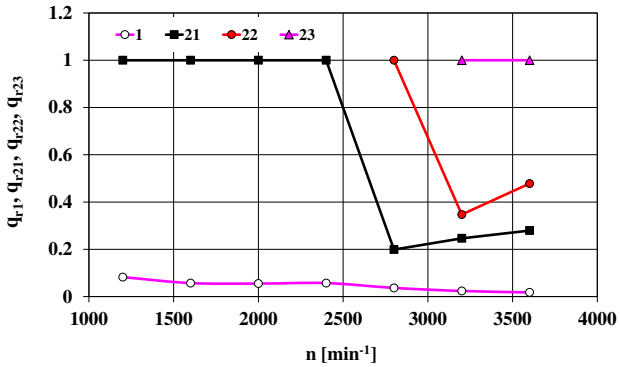

Figure 22

The relative unit heat – q_r (1 – in the first injection, 21 – in the first part of the second injection, 22 – in the second part of the second injection, 23 – in the third part of the second injection) as a dependence on the engine rotational speed

The unit heat being released in the first injection is almost constant over the entire useful range of rotational speed, while in the second injection the unit heat increases with the increasing rotational speed – consequently, the amount of fuel dosed by the supply system is an increasing function of the rotational speed.

Conclusions

The dynamics of the processes associated with the combustion of fuel in the cylinders of compression-ignition engines with multiple fuel injection is determined primarily by the flow rate of fuel injected into the cylinders in the individual injections. With Common Rail supply systems, the dose of fuel injected into the cylinders is primarily determined by the opening time of the injectors. It is significant to the analyzed fuel supply system that the increasing flow of fuel dosed, with increasing rotational speed, is determined mainly by the second injection, while the initiation of the combustion process is propagated by an earlier start of the injection, with increasing rotational speed.

Both the autoignition delay angle of the first fuel dose and the ignition-delay angle of the second fuel dose are increasing functions of rotational speed, which corresponds approximately to the constant value of the delay time of the autoignition and ignition of the second fuel dose. At the same time, the ignition-delay angle of the second fuel dose is substantially smaller than the autoignition delay – this is the result of the different nature of phenomena of the combustion initiation in both cases in question.

The research work carried out proved that the angle of the maximum value of the unit heat-release rate is weakly correlated with the angle of the maximum value of the control signal in each phase of ignition, although the rate of heat release is determined by the flow of fuel supplied to the cylinder. This is confirmed by the

results of correlation analyses of the processes of the injector-opening control voltage and unit heat-release rate. The angle of the delay to the unit heat-release rate, in respect of the signal of the control voltage of the injector opening, is approximately 10–40 deg and is an increasing function of the rotational speed. The time of delay to the unit heat-release rate, in respect of the signal of the control voltage of the injector opening, is weakly dependent on the rotational speed, as it depends mainly on the kinetics of pre-ignition reactions.

The research work carried out proved that the maximum unit heat-release rate is much higher in the second injection than in the first injection. The unit heat being released in the first injection is almost constant over the entire useful range of the rotational speed, while in the second injection the unit heat increases with increasing rotational speed.

Research into the heat release dynamics in the cylinder of compression-ignition engines with the multiple injection of fuel allows to gain knowledge about the processes occurring in engine cylinders, and – consequently – to rationally assess the impact of fuel supply on engine performance.

References

[1] Abbaszadehmosayebi, G.: Diesel Engine Heat Release Analysis by Using Newly Defined Dimensionless Parameters. PhD thesis. Brunel University, London 2014

[2] Ambrozik, A., Orliński, P., Orliński, S.: Influence of diesel engine fuelling with diferent fuels on self-ignation delay in aspect of ecology. Eksploatacja i Niezawodność – Maintenance and Reliability, 2003, 19, 50-55

[3] Atzler, F., Kastner, O., Rotondi, R., Weigand, A.: Multiple injection and rate shaping Part 1: emissions reduction in passenger car Diesel engines. SAE Technical Paper 2009-24-0004, 2009

[4] AVL: Testing solutions. Measure and control. Single-cylinder research engines & Compact test bed. [online] Available at: https://www.avl.com/-/single-cylinder-research-engines-compact-test-bed [Accessed 20.02.2019]

[5] Bendat, J. S., Piersol, A. G.: Random Data: Analysis & Measurement Procedures. Wiley, New York, 1966

[6] Chłopek, Z., Jagiełło, S., Juwa, S., Skrzek, T.: Comparative examination of performance characteristics of an IC engine fuelled with diesel oil and rape methyl esters. The Archives of Automotive Engineering – Archiwum Motoryzacji, 2016, 74, 19-32

[7] Chłopek, Z.: On selected methods of numerical differentiation on the example of cylindrical pressure course differentiation. Journal of KONES, 2001, 8, 147-153

[8] Chłopek, Z.: The subject of task explicitness in numerical differentiation. Journal of KONES, 2001, 8, 302-311

[9] d'Ambrosio, S., Ferrari, A.: Potential of multiple injection strategies implementing the after shot and optimized with the design of experiments procedure to improve diesel engine emissions and performance. Applied Energy, 2015, 155, 933-946

[10] Di Iorio, S., Mancaruso, E., Vaglieco, B. M.: Characterization of Soot Particles Produced in a Transparent Research CR DI Diesel Engine Operating with Conventional and Advanced Combustion Strategies. Aerosol Science and Technology, 2012, 46, 272-286

[11] Gyujin, K., Moon, S., Lee, S., Min, K.: Numerical Analysis of the Combustion and Emission Characteristics of Diesel Engines with Multiple Injection Strategies Using a Modified 2-D Flamelet Model. Energies, 2017, 10(9), 1292

[12] Han, Z., Uludogan, A., Hampson, G. J., Reitz, R. D.: Mechanism of soot and NOx emission reduction using multiple-injection in a diesel engine, SAE Technical Paper 960633, 1996

[13] Heywood, J. B.: Internal Combustion Engine Fundamentals. McGraw-Hill, New York, 1989

[14] Khandal, S. V., Banapurmath, N. R., Gaitonde, V. N., Hiremath, S. S.: Paradigm shift from mechanical direct injection diesel engines to advanced injection strategies of diesel homogeneous charge compression ignition (HCCI) engines – A comprehensive review. Renewable & Sustainable Energy Reviews, 2017, 70, 369-384

[15] Lamas, M. I., de Dios Rodríguez, J., Castro-Santos, L., Carral, L. M.: Effect of multiple injection strategies on emissions and performance in the Wärtsilä 6L 46 marine engine. A numerical approach. Journal of Cleaner Production, 2019, 206, 1-10

[16] Lim., J., Lee, S., Min, K.: Combustion modeling of split injection in HSDI diesel engines. Combustion Science & Technology, 2010, 183, 180-201

[17] Liu, H -P., Strank, S. M., Werst, M. D., Hebner, R. E., Osara, J.: Combustion emissions modeling and testing of conventional diesel fuel. In: Proc. ASME 2010 4th International Conference on Energy Sustainability. Phoenix: CEM Publications, 2010

[18] Men, Y., Haskara, I., Zhu, G.: Multi-zone reaction-based modeling of combustion for multiple-injection diesel engines. International Journal of Engine Research, 2018, 1468-0874

[19] Merker, G., Schwarz, C., Stiesch, G., Otto, F.: Simulating Combustion. Springer-Verlag, Berlin, Heidelberg, 2006

[20] Oppenheim, A. V., Schafer, R. W.: Digital Signal Processing. Pearson, New Jersey, 1975

[21] Pearson, K.: On the Theory of Contingency and Its Relation to Association and Normal Correlation, Dulau and Co., London, 1904

[22] Rakowski, S., Eckert, P., Witt, A.: Engine Combustion. In: Merker, G., Schwarz, C., Teichmann, R. (Eds), Combustion Engines Development. Springer, Berlin, Heidelberg, 2012, Chap. 4, 119-168

[23] Reitz, R. D.: Controlling D. I. diesel engine emissions using multiple injections and EGR. Combustion Science & Technology, 1998, 138, 257-278

[24] Samoilenko, D., Prokhorenko, A., Marchenko, A.: Simulation of processes in variable geometry turbine of high speed diesel engine. In: Proceedings of the 19th International Conference on Transport Means 2015, Kaunas, 22-23.10.2015

[25] Tauzia, X., Maiboom, A., Chesse, P., Thouvenel, N.: A new phenomenological heat release model for thermodynamical simulation of modern turbocharged heavy duty diesel engines, Applied Thermal Engineering, 2006, 26, 1851-1857

Test Result Utilization in the Development Process of High Voltage SF6 Circuit-Breaker

Zsolt Mitrik, Péter Kádár

Óbuda University, Kálmán Kandó Faculty of Electrical Engineering,
Power System Department, Alternative Eneregy Sources Laboratory
Bécsi út 94-96, H-1034 Budapest, Hungary
mitrik.zsolt@kvk.uni-obuda.hu; kadar.peter@kvk.uni-obuda.hu

Abstract: During the development tests of high voltagecircuit-breakers (HVCB), the manufacturers obtain a large amount of information regarding the behavior and parameters of the equipment. Such valuable conclusions can be deduced from processing, analyzing and evaluating them by scientific methods that are appropriate for developing a new decision-supporting model. The Author describes the methodology of the development process of such part of the new decision-supporting model (Circuit Breaker Design algorithm CBDa) which systematizes and evaluates the information from tests as results and illustrates the possible ways of usage.

Keywords: HVCB; 3rd generation SF6 circuit-breaker; CBDa model

1 Introduction

High-voltage circuit-breakers are an integral part of the electricity system. Their task is to close and open the circuits in normal operational and outside of normal operational circumstances. The circuit-breakers are the only short circuit-protection devices (SCPD) in the high voltage networks. Accordingly, reliable operation is critical in terms of reliability of the entire system. It is necessary to extinguish the arc between the contacts that appears during the opening process, within the shortest time, even several times in succession at the breaking of the short-circuit currents. The arc chambers built into the circuit-breaker perform this task [1]. The Author analyses those high voltage circuit-breakers that use SF6 gas as arc extinguisher medium, which is the same as the insulator medium. The Author investigates the third generation, auto-puffer high voltage circuit-breakers, as those are the most common and uses the most modern technology. The IEC 62271-100:2008/AMD2:2017/COR1:2018 [2] refers to high voltage circuit-breakers and describes what types of breaking tests certify that the tested circuit-breaker is able to perform its tasks under any network conditions and would be able to break the circuit safely. During the design of the circuit-breakers, the

designers follow two main directions. Firstly, they perform a computer simulation in which they analyze the variation of the generated gas flow (speed, density, pressure change) and the electrical field strengthen the developed geometric circumstances [3]. On the other hand, based on the results of the simulations, test equipment is manufactured on which the designers carry out the tests in order to verify the correctness of the calculations. The Author created and introduced to the science a new Circuit Breaker Design algorithm (CBDa) model [4] as a result of the Author's scientific work which is suitable for modeling the SF6 high voltage circuit-breaker development process. Figure 1 shows the CBDa model's main steps:

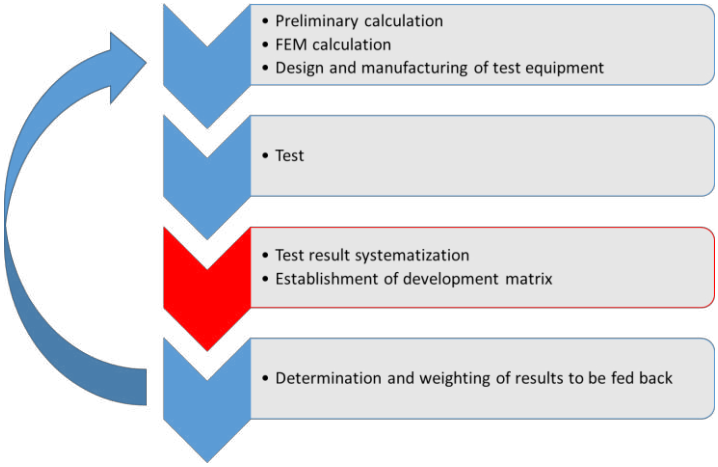

Figure 1
Simplified flow chart of CBDa model

This article describes the third step of Figure 1 in detail. In the following chapters, the Author presents the evaluation of results and information from the completed tests and the utilization of them. The fourth step of the model (Figure 1) is the determination and weighting of results to be fed back [5]. The author will describe the feedback method in then extarticle. It will also include the methodology of weighting and how all these increase the accuracy of preliminary calculations and simulations.

2 Theory and Process of Measurement

The measurement starts with the determination of the scope of the measurement. The first step is to determine what one would find out with the measurement. The second step of the measurement process is the collection of all prior information and knowledge. It is investigated what other researchers used what variables and

which methods and what results they reached. After all these, one plans the measurement process: what is measured, which equipment is used for the measure, what protocol will apply for the tests. At the next step, the measurement equipment is established and the measurement instrument is checked [6]. The actual measurement/test is processed after this and if needed, correction can be set up in the measurement process. The final step of the measurement process is the evaluation of the test results. [7]

2.1 Application of the Measurement Theory

The measurement theory [8] is a scientific tool used during the measurement process: measurement planning, data collection, measurement and evaluation of results. It can be applied and performed step-by-step: (Fig. 2).

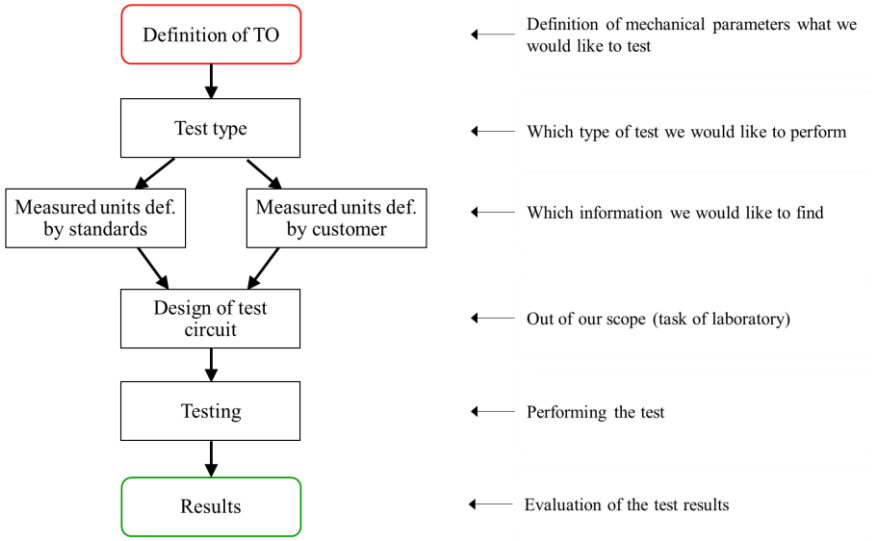

Figure 2
Circuit-breaker test flow chart

Figure 2 shows the test process. The first step is the determination of the geometry to be tested, i.e. the assembly of the test circuit-breaker (Test object – TO). In the next step it is determined in accordance with the CBDa model [4] and the standards (2.2) which test would be performed. Then, the researcher defines the measurable physical quantities. The standards prescribe part of these (typically the electrical parameters of the test circuit). However, such additional parameters can be determined that can be used at a later stage also for feedback. The test laboratory can design their own test circuit once the measurement circuit and supplementary parameters are determined. The tests can be performed after these steps. The Author describes the evaluation of the results in the last chapter (4).

2.2 Standard Environment of Synthetic Tests

The IEC 62271-100:2008/AMD2:2017/COR1:2018 [2] applies to the high voltage circuit-breakers. The standard prescribes the carry-out of all such test that needs for the certification that the high voltage circuit-breaker can be built into the system, and also for the proof of the proper operation in the system and for the proof of the lifetime. In this article, the author deals only with the breaking tests and the evaluation of such tests. The fulfillment of the following tests prove that the circuit-breaker operates properly under any circuit circumstances:

1. Basic short-circuit tests (T100s)
2. Basic short-circuit tests (T100a)
3. Basic short-circuit tests (T60)
4. Basic short-circuit tests (T30)
5. Basic short-circuit tests (T10)
6. Short-line fault test (L90)
7. Short-line fault test (L75)
8. Out-of-phase making and breaking test (OP2)
9. Capacitive current switching test (BC2) Making / Breaking
10. Capacitive current switching test (BC1)
11. Double-earth fault test (DEF)

One part of the above tests are opening tests (O), other parts are closing and opening tests (CO) and there are tests to simulate the opening – closing– opening operations (O-0,3s-CO) with that the fast reclose automation is simulated.

Certain tests must be performed with different arcing times. The minimum arcing time in this context is the shortest time under which the circuit-breaker in the set circuit conditions can interrupt the circuit. Some tests are performed with maximum arcing time, others with medium arcing time and there are combinations of these. The standards prescribe that the minimum arcing time must be found first and all other test times must be calculated from this value.

2.3 The Process of Measurement

Due to the high energy need of the above-described tests, it is not possible to perform the test on the electric network. So-called synthetic circuits are used for these tests. Figure 3 shows a sample synthetic circuit. [9]

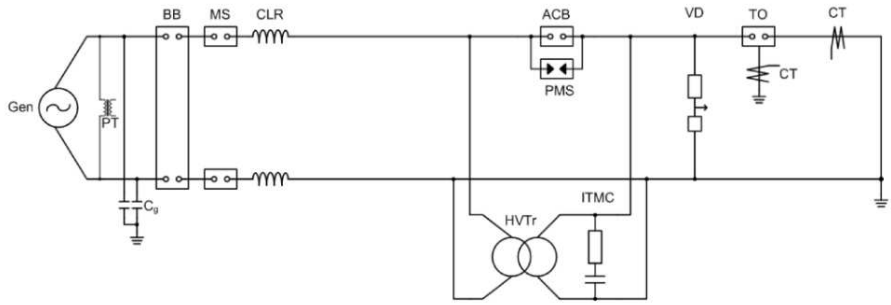

Figure 3
Test circuit for basic short-circuit tests (T100s(a)) [10]

The Figure 3 shows the simplified schematic connection drawing of a synthetic circuit that is suitable for T100s (a) test. Other, similar circuits are assigned to each test types. The characteristics of the network – the test laboratory wants to simulate – can be reached by the fine-tuning of the built-in inductive and capacitive elements. [11]

On Figure 3 there are the following marks: Gen - generator, PT – potential transformer, Cg – TRV tuning capacitor, BB – back-up circuit-breaker, MS – making switch, CLR – current limiting reactor, ACB – auxiliary circuit-breaker, PMS – Plasma making switch, VD – voltage divider, TO – test object, CT – current transformer, HVTr – high voltage transformer, ITMC – Initial transient making current circuit.

The standards prescribe that the minimum arcing time must be found first and all other test times must be calculated from this value. However, this means that tests must be carried out on the circuit-breaker in a way that the arcing time is decreased by milliseconds until the circuit-breaker cannot interrupt the circuit. Accordingly, a circuit-breaker carries out more breaks even before the real tests. This fact is important as certain assembled parts are damaged or worn with each breaking. This worn may affect the arcing contacts and the material (Teflon) of the gas deflector components (e.g. main nozzles) around the arcing place which evaporates due to the extreme heat. As a consequence of the evaporation of this material, the geometry of the nozzles changes and this decreases the efficiency of the arc extinguish. Figure 4 shows the simplified assembly of the arcing chamber of an SF6 circuit-breaker.

Figure 4 shows an arcing chamber of the circuit-breaker in which 1 is the gas deflector, 2 and 6 are arcing contacts, 3 – arc formation place, 4, 5 – main contacts.

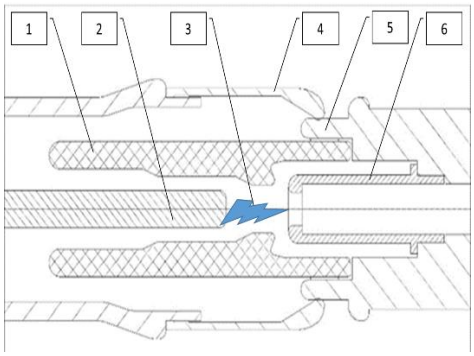

Figure4
Simplified assembly of circuit-breaker arcing chamber (own figure)

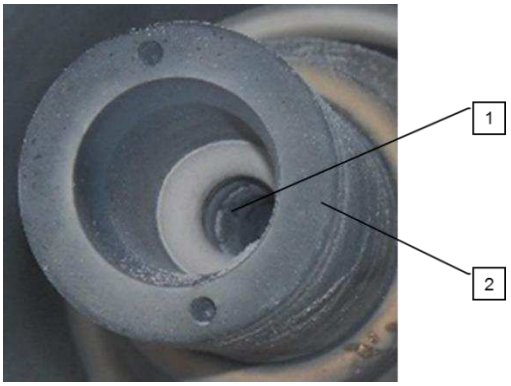

Figure 5
Part of used arcing chamber (own photo)

Figure 5is a picture about a part of a used arcing chamber. It can be observed that the arcing sleeve (1) and the main nozzle (2) melted and evaporated due to the heat.

During the developmental process, several geometries is tested. These typically relate to the volume of the thermal and compression chambers, the establishment of that geometry that determines the gas flow channel. Generally, the testing process starts with tests that load the circuit-breaker the most. These are the 100% breaking and the L90 test duties. At the same, time it is necessary to consider that due to the operational principle of third-generation (auto-puffer) SF6 circuit-breakers, the range of medium short-circuit currents, test duty (T60), represents a significant strain [12]. If the first test on the geometry established by the computer simulation is successful then the next test is carried out on the same geometry. During the test a large amount of information is obtained on the characteristics of the circuit-breaker. The evaluation and feedback of this information into the development process may increase the accuracy and effectiveness of the development.

3 The Measurement and the Organization of Measurement Results

The large amount of data generated during the development process classified into three groups: mechanical parameters (3.1), test parameters (3.2) and the ordering of measurement results (3.3).The aggregated matrix, as the result of the data in the three groups, is presented in the chapter 3.4. [13]

3.1 Ordering of Mechanical Parameters

Several parts are assembled in the arcing chambers of third-generation SF6 circuit-breakers. The Author – based on prior experiences – determined sixteen different mechanical parameters (examined geometric shape) which have major impact on the arc extinguish process. These mechanical parameters are marked as M_1-M_{16}.

During the tests the parameters are modified continuously and are summarized in the matrix (1):

$$A_{TM} = \begin{pmatrix} M_{1,1} & M_{1,2} & \dots & M_{1,16} \\ M_{2,1} & M_{2,2} & \dots & M_{2,16} \\ \dots & \dots & \dots & \dots \\ M_{n,1} & M_{n,2} & \dots & M_{n,16} \end{pmatrix} \quad (1)$$

where the matrix columns indicate a mechanical parameter and each line corresponds to a test configuration, shape combination.

Mechanical parameter		M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12	M13	M14	M15	M16
Test number	37	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	38	2*)	175	24	33	---	12 / 19	42 / 67	20	R11	3	normal	125	NO	320	500	37
	39	2*)	175	24	33	---	3 / 19	36 / 67	14	R11	3	normal	125	866530-I	320	500	37
	40	2*)	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	41	5	187	24	33	---	12 / 19	36 / 79	14	R11	1	normal	145	866530-I	320	500	37
	42	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
	43	4	187	24	34,4	---	3 / 19	36 / 79	14	R11	1	normal	145	866530-I	320	550	37

Figure 6

Mechanical data in table format

Lines are taken on till the given test type becomes successful.

3.2 Ordering of Test Parameters

According to the above, the test parameters are also included in a matrix (2). E_1 indicates the test type (Chapter 2 standard environment), E_2 – E_9 are the measured electric and mechanical values. Even though the standards determine the value of the electric parameters (e.g. transient recovery voltage (TRV) value, rate of decrease of the specified short-circuit current (di/dt), etc.) these can differ from the nominal value due to the standard deviation of the synthetic circuit. Therefore, the

author considers these values as the measured value. The matrix (2) also contains mechanical values that are typical for that test (e.g. set gas pressure, operational speed and main time of circuit-breaker, etc.):

$$A_{TE} = \begin{pmatrix} E_{1,1} & E_{1,2} & \dots & E_{1,8} \\ E_{2,1} & E_{2,2} & \dots & E_{2,8} \\ \dots & \dots & \dots & \dots \\ E_{n,1} & E_{n,2} & \dots & E_{n,8} \end{pmatrix} \tag{2}$$

where the matrix columns indicate a test parameter. Similarly to the mechanical parameters, the lines are taken on till the test type becomes successful.

Electrical parameter		E1		E2	E3	E4	E5	E6	E7	E8	E9
Test number	37	L90	AQF0507	5,98	6	22,07	124	35,6	15,93	---	169
	38	L90	AQF0606	6	6	17,4	125	35,74	16,34	---	
	39	L90	AQF1104	6	6	21,44	125	35,81	16,4	---	170
	40	L90	AQF1106	6	6	24,9	125	35,7	16,35	---	
	41	L90	AQF2809	5,79	6	14,64	123	36,3	15,8	---	161
	42	L90	AQF2819	5,8	6	21,71	125	36,06	16,39	---	
	43	L90	AQG0515	6	6	21,7	124	35,7	16,19	---	165

Figure 7
Test parameters in table format

Figure 7 shows an example of the electric parameters of L90 tests. Similar tables are generated to collect the results of each test types that are specified in point 2.2.

3.3 Ordering of Measurement Results

$R_1 – R_{11}$ indicates the measurement results. R_1may have two different values, 1 – if the test was successful and 0 – if the test failed. $R_2 – R_{11}$ indicates such kinds of parameters as, e.g. first and second peak value of the pressure in the thermal chamber. This data is also integrated into a matrix (3):

$$A_{TR} = \begin{pmatrix} R_{1,1} & R_{1,2} & \dots & R_{1,11} \\ R_{2,1} & R_{2,2} & \dots & R_{2,11} \\ \dots & \dots & \dots & \dots \\ R_{n,1} = 1 & R_{n,2} & \dots & R_{n,11} \end{pmatrix} \tag{3}$$

As before, each line in the matrix includes the results of a new test.

Mesurement results		R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11
Test number	37	OK/NO	121	3	X	X *)	---	---	---	---	---	*) 82,5 kV, after 6ms
	38	NO	113	---	---	---	9,42	---	13,03	10,7	---	---
	39	OK	---	---	---	---	34,2	---	45,7	33	---	---
	40	NO*)	---	---	---	---	30	---	40,4	28,3	---	*) membrane breach
	41	OK	---	---	---	---	27,71	19,9	---	---	---	---
	42	NO	---	---	---	---	24,26	18,5	24,16	16,36	---	---
	43	OK/NO	105	16,3	---	---	43,17	35,9	39,27	30,3	---	---

Figure 8
Measurement results in table format

Figure 7 and Figure 8 show details of the test parameters and the test results tables. The missing values are such values that are typically non-test data, or data not registered due to measurement circuit failure.

3.4 The Aggregated Matrix

Description of the development measurement series

1. M_1-M_{16}defines the mechanical characteristics of the equipment. The researcher runs a standard described test type, e.g. L90. This is indicated as T1. It is almost certain that the first test will fail, therefore, R_1will be 0.
2. M_1-M_{16} values are modified until the test becomes successful. At this time R_1will be 1.
3. Once R_1=1, the measurements of next test type T_2 start. It is also almost certain that the first test is unsuccessful. M_1-M_{16} values are again modified until R_1 becomes 1.
4. Then the researcher returns to T_1. Based on the prior measurements, it is clear that what kind of modification of which element of M_1-M_{16} what effect it has on the result. M_1-M_{16} is modified in such a way that R_1 = 1, in both T_1 and T_2. This can be reached from several tests.
5. If it is reached that for the same M_1-M_{16}, R_1 = 1 both at T_1 and T_2, then T_3 can start.
6. This process is continued for all T_x until R_1 = 1 using the same M_1-M_{16} parameters.
7. R_2…R_{11}results are used for the determination of the next M_1-M_{16} independently from the fact of whether R_1 value is 0 or 1.

Aggregating the matrixes of the mechanical parameters, of the test parameter and the matrix of the results, a new matrix is received that is the aggregated matrix of one test type:

$$A_{T1} = \begin{pmatrix} M_{1,1} & M_{1,2} & \dots & M_{1,16} & E_{1,1} & \dots & E_{1,8} & R_{1,1} & R_{1,2} & \dots & R_{1,11} \\ M_{2,1} & M_{2,2} & \dots & M_{2,16} & E_{2,1} & \dots & E_{2,8} & R_{2,1} & R_{2,2} & \dots & R_{2,11} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ M_{n,1} & M_{n,2} & \dots & M_{n,16} & E_{n,1} & \dots & E_{n,8} & R_{n,1} = 1 & R_{n,2} & \dots & R_{n,11} \end{pmatrix} \quad (4)$$

Figure 9 shows the application of A_{TM}, A_{TE}, A_{TR} and the aggregated matrix in the previously published CBDa model. The determination of the geometry in the CBDa model corresponds to a new line of A_{TM} matrix. At the same way, the new measured parameters and the measurement results in the tests would increase A_{TE}, A_{TR} matrixes. The aggregated matrix represents the amount of information to feedback.

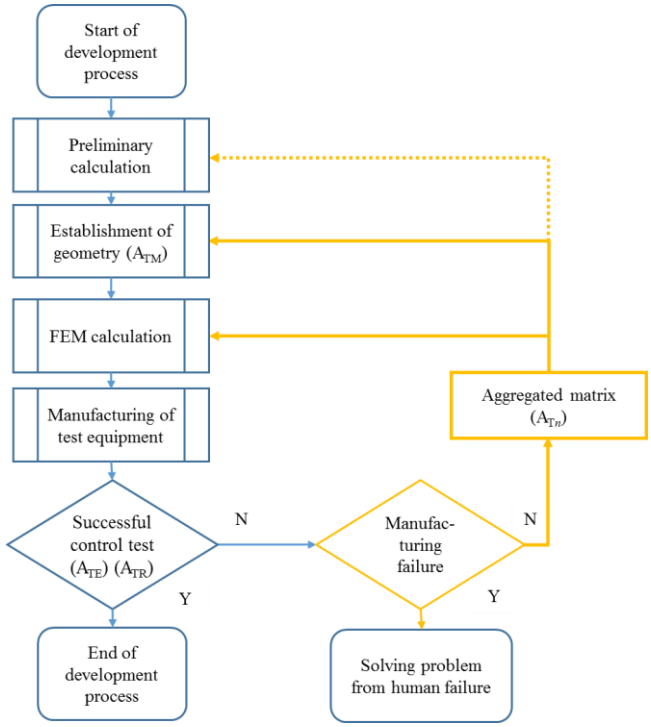

Figure 9

Development process with feedbacks [4]

Figure 9 shows the development process and the integration of the aggregated matrix in the feedback processes. The Authors described the development process in detail in their previously published article. [4]

4 Evaluation of Feedback Results

The author, former leader of R&D team, in Switchgear Department, at Ganz Transelektro Electric Co. took part in several tests performed at VEIKI-VNL Electric Large Laboratories Ltd. Figure 10 and Figure 11 show two test results carried out in this laboratory. The oscillograms contain a completed test result, which is the test of the same circuit-breaker on the same test type, but with different arcing time:

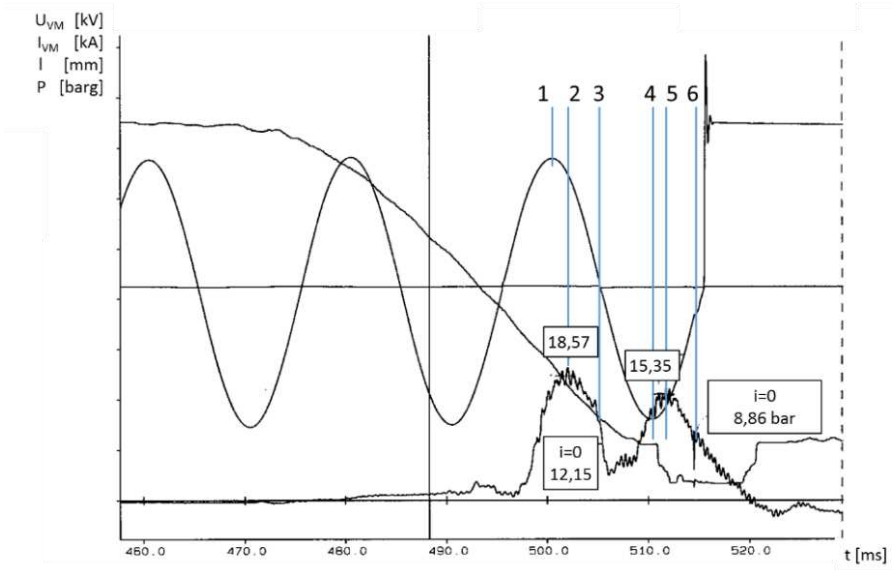

Figure 10
Breaking process with T100s maximum arcing time

Figure 10 shows the current flowing through the poles of the circuit breaker, the voltage between the poles after the current interruption process, the time-travel diagram and the gas pressure evolution in the thermal chamber.

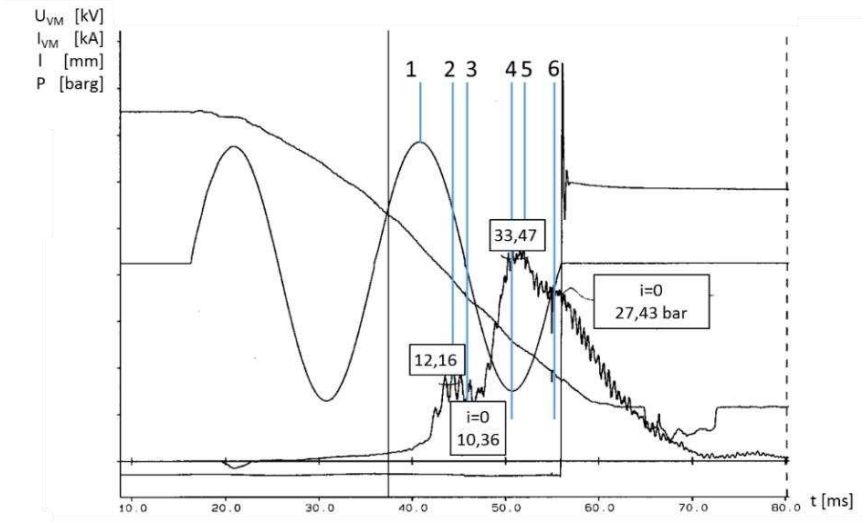

Figure 11
Breaking process with T100s minimum arcing time

Figures 10 and 11 show the operation of the same circuit-breaker during a T100s breaking process. Figure 10 shows the maximum while Figure 11 shows the minimum arcing time. The current through the tested circuit-breaker, the transient recovery voltage, travel-time characteristic, and the pressure evolution in the thermal chamber show in the diagrams. Points 1 and 4 show the current's first and second peak values after contact separation. The pressure peak values also appear (points 2 and 5). The time shift of pressure peak values in relation to the current peak values is the consequence of the arc thermal inertia. Point 3 indicates the current's zero before breaking. Point 6 shows the start of the operation of the current injection circuit [14], [15].

The diagrams (Figure 10, Figure 11) highlight two substantial facts:

- First is that the pressure peak value arises after the current peak value. This is the effect of the arc inertia.
- Second is that the thermal chamber pressure at the maximum arcing time is lower (8,86 barg on Figure 10) then in case of the minimum arcing time (27,43 barg on Figure 11).

Due to the facts described in Chapter 3, the Author transforms the engineering task into a mathematical task with the use of the feedback method discussed in Chapter 4. Accordingly, the Author studied the $M_{n,1}..M_{n,16}$ values where for any E_1 $R_1 = 1$. On the other hand, the $M_{n,1}..M_{n,16}$ values were also examined for the case where p provides an ideal distribution.

Conclusions

With the assistance of measurement theory and considering the standards' specifications, in this article the Author determined those typical mechanical parameters of the arcing chamber of circuit-breaker which have the greatest influence on the circuit-breaker's operation. These are the bases on which A_{TM} matrix consisting of the mechanical parameters is set. The modification of A_{TM} matrix resulted in the test parameter matrix including E_1E_9. The Author denominated 11 result parameters that used during the feedback process. From all of this data, the A_{TR} result matrix is set and with the connection and aggregation of all matrixes a final matrix is set. This is the aggregated circuit-breaker development matrix. In the article, the interpretation of one test results is shown based on the oscillogram of two arcing time of T100s test.

In the following article the changes in the geometry will be associated with the pressure wave peak values. The Author will weigh the A_{TR} mechanical parameters to be fed back by using Fuzzy logic. This is necessary as the FEM calculation will be fine-tuned with a feedback appropriate to CBDa model and these results will be used in the establishment of the new geometry. The Author will present the multidimensional matrix established from the aggregated matrixes used for the different tests.

Acknowledgement

The research was supported by AD&TE research group of Faculty of Electrical Engineering of Óbuda University.

References

[1] Dr. Madarász György Attila: "Breackers. Part 1: High voltage circuit-breaker" Elektrotechnika 2016/1-2, pp. 15-19

[2] International Electrotechnical Commission, https://webstore.iec.ch/publication/62278, 2019.03.02, 10:24

[3] István Vajda, Alexander Glazyrin, Irina Ustinova, Evgeniy Bolovin: "Influence of Design Methods a Discrete Model of Separately Excited DC Motor on Parameters Estimation" Acta Polytechnica Hungarica Vol. 15, No. 6, 2018, pp. 219-233

[4] Zsolt, Mitrik; Péter, Kádár "System theory approach of SF6 high voltage circuit breaker development process" SAMI 2018 : IEEE 16th World Symposium on Applied Machine Intelligence and Informatics: Dedicated to the Memory of Pioneer of Robotics Antal (Tony) K. Bejczy: proceedings Seattle (WA), United States of America : IEEE (2018) pp. 137-142, 6 p.

[5] László Horváth, Imre J. Rudas: "Information Content Driven Model for Virtual Engineering Space" Acta Polytechnica Hungarica, Vol. 15, No. 2, 2018, pp. 7-32

[6] Jana Moravčíková, Peter Pokorný: "The Influence of Machine-Part Measuring Strategies for Coordinate Measuring Devices on the Precision of the Measured Values" Acta Polytechnica Hungarica, Vol. 15, No. 6, 2018, pp. 7-26

[7] Dr. Kopper Bence, Methodology of scientific research, tf.hu/wp-content/uploads/.../Tudományos-kutatásmódszertan.pptx, Viewing date: 2019.04.19, 16:51

[8] Halász-Huba: Technical measurments. Technical University Publishing Office 2003. ISBN 96342074]

[9] Kourosh Mousavi Takami, Erik Dahlquist "Modeling and simulation of short circuit current and trv to Develop a synthetic test system for circuit breakers" The 55th Conference on Simulation and Modelling (SIMS 55) 21-22 October, 2014, Aalborg, Denmark, Page: 285 -

[10] Park, Byung-Rak, Korea Electrotechnology Research Institute: "Type test certificate" 2011.04.22. Doc. No.: 2011TC00238

[11] Z. Conka, M. Kolcun, G. Morva, „Utilizing of Phase Shift Transformer for increasing of Total Transfer Capacity" In: Acta Polytechnica Hungarica, Vol. 13, Issue:5, pp. 27-37, 2016, DOI: 10.12700/APH.13.5.2016.5.2, ISSN: 1785-8860

[12] Zsolt Mitrik, "Operation of third generation sulphur- hexafluoride circuit breakers in medium zone of short circuit currents", 32nd Kandó Conference 17 of November 2016, ISBN 978-963-7158-07-0

[13] R. Andoga, L. Madarász, L. Főző, T. Lazar, V. Gaspar, "Innovative Approaches in Modeling, Control and Diagnostics of Small Turbojet Engines" In: Acta Polytechnica Hungarica, Vol. 10, Issue:5, pp. 27-37, 2013, DOI: 10.12700/APH.10.05.2013.5.6, ISSN: 1785-8860

[14] Rümpler Ch,-Narayanan V.R.T.: Arc Modelling Challenges, Plasma Physics and Technology, 2015, Vol. 2, No. 3, pp. 261-270

[15] Niu Ch. at al.: Simulation and Experimental Analysis of Arc Motion Characteristics in Arc Circuit Breaker, Plasma Physics and Technology, 2016, Vol. 18, No. 3, p

Camera-based On-Road Detections for the Visually Impaired

Judith Jakob[1,2], József Tick[3]

[1] Óbuda University Budapest, Doctoral School of Applied Informatics and Applied Mathematics, Bécsi út 96/b, 1034 Budapest, Hungary

[2] Furtwangen University of Applied Sciences, Faculty of Computer Science, Robert-Gerwig-Platz 1, 78120 Furtwangen, Germany
e-mail: jaju@hs-furtwangen.de

[3] Óbuda University Budapest, Institute of Applied Informatics, John von Neumann Faculty of Informatics, Bécsi út 96/b, 1034 Budapest, Hungary
e-mail: tick@uni-obuda.hu

Abstract: Herein, we research the possibilities of assisting visually impaired pedestrians moving in traffic situations by using camera-based detection of relevant objects in their immediate surroundings. Therefore, we use and adapt algorithms from the field of driver assistance. We present a road background segmentation based on watersheds, whose results are used as input for the presented crosswalk and lane detection algorithms. The crosswalk detection is based on the application of two 1D mean filters and the lane detection on local computations of the EDF (Edge Distribution Function). In our evaluation, the described algorithms achieved good hit rates of 99.87% (road segmentation), 98.64% (crosswalk detection), and 97.89% (lane detection).

Keywords: ADAS; Driver Assistance; Visually Impaired; Road Segmentation; Crosswalk Detection; Lane Detection; Computer Vision

1 Introduction

According to [6], in 2015, worldwide 36 million people were estimated to be blind and 216.6 million people were estimated to have a moderate to severe visual impairment. Since 1990, the numbers of both groups have increased: in comparison to 1990, there was a rise of 17.5% in blind people and 35.5% in people with moderate to severe visual impairment.

In their study from 2001, Duckett and Pratt underline the importance of mobility for the visually impaired: “The lack of adequate transport was described as resulting in many visually impaired people living in isolation. Transport was felt

to be the key to visually impaired people fulfilling their potential and playing an active role in society," [10].

Therefore, we research possibilities of assisting visually impaired people by using camera-based object detection in traffic situations in order to increase their independent mobility. As the amount of research in the field of Advanced Driver Assistance Systems (ADAS) is much higher in comparison with Assistive Systems for the Visually Impaired (ASVI), our research focuses on transferring algorithms and concepts from ADAS to ASVI.

1.1 Previous Work

In [19], we motivate the need for a transfer concept from ADAS to ASVI and present a plan for an ASVI consisting of a smartphone app and a cloud service. A camera as well as earphones to provide text-to-speech output are connected to the smartphone; expensive image processing calculations are exported to the cloud service. Although the system will not be built in the near future, we keep this system in mind as reference system during our research. Furthermore, [19] contains first findings regarding the transfer concept and a short description of a preliminary version of the crosswalk detection algorithm that we describe in detail in this article.

In order to gather the needs of visually impaired people in traffic situations, we conducted qualitative interviews with experts and members of the target group using Witzel's method of problem-centered interviews [34]. The results of the expert interviews are presented in [18] and [20]; a common evaluation of both interview types is not published yet. The evaluation of the interviews is presented in the form of tables, adapted from Sommerville's software engineering book [30], describing the six identified traffic scenarios: (1) General orientation, (2) navigating to an address, (3) crossing a road, (4) obstacle avoidance, (5) boarding a bus, and (6) at the train station. For each traffic scenario, we recorded the vision use cases that could help visually impaired people in the according scenario. We then formed the overlap with vision use cases that are of interest in both considered fields, ADAS and ASVI. These are the use cases, we have to consider in our research: (1) Lane detection, (2) crosswalk detection, (3) traffic sign detection, (4) traffic light (state) detection, (5) (driving) vehicle detection, (6) obstacle detection, and (7) bicycle detection. Additionally, we described a preliminary version of the lane detection algorithm presented explicitly in this article in [20].

For the evaluation of the algorithms that are developed in the course of our research, comparable sequences from driver and pedestrian perspective covering the seven identified use cases are needed. While there are many publicly available data sets from driver perspective, e.g. the KITTI data set [14] and the German Traffic Sign Detection Benchmark [16], there are few data sets from pedestrian perspective and no data sets with comparable material from both perspectives.

Therefore, we created the CoPeD (Comparable Pedestrian Driver) data set for traffic scenarios [21]. It is licensed under the Creative Commons Attribution 4.0 International License[1] and hosted publicly[2]. The data set is divided into four categories: (1) Lanes, (2) crossings (incl. traffic lights and crosswalks), (3) obstacles (incl. vehicles and bicycles), and (4) traffic signs with relevance to pedestrians.

1.2 Presented Work

The next step towards a transfer concept from ADAS to ASVI is the examination of the seven identified use cases with relevance in both fields concerning their possibilities of adaptation from ADAS to ASVI.

The additional consideration of a road background segmentation makes it possible to run some detection algorithms on only a subset of the original image. Crosswalk and lane detection can be executed on the road image, whereas traffic sign and traffic light detection are carried out on the background part of the image. Obstacles, including vehicles and bikes, can occur in both parts of the image and are therefore run on the complete image. Figure 1 illustrates the procedure. This article focuses on the middle path of the diagram in Figure 1, meaning that we propose a road background segmentation and solutions for the "on-road" detections crosswalks and lanes. Thereby, the segmentation result is used as a preprocessing step for the different detection algorithms. This means that when integrated into an assistive system, the segmentation result is not communicated to the user unlike the result of the requested detection algorithm.

The following chapter first gives an overview concerning related work for the topics camera-based ASVI as well as road background segmentation, crosswalk detection, and lane detection for ADAS and ASVI. Afterwards, we present our proposed ASVI solutions for these topics, each based on procedures from ADAS. The proposed algorithms are then evaluated on sequences from the CoPeD data set [21]. Finally, Chapter 4 outlines our future work.

2 Related Work

This chapter first gives an overview of camera-based assistance for the visually impaired. Then, related work of the three topics that have to be considered, namely road background segmentation, crosswalk, and lane detection, is cited. For each topic, we explain which ADAS procedures we chose for the adaptations presented in the following Chapter 3.

[1] https://creativecommons.org/licenses/by/4.0/
[2] http://dataset.informatik.hs-furtwangen.de/

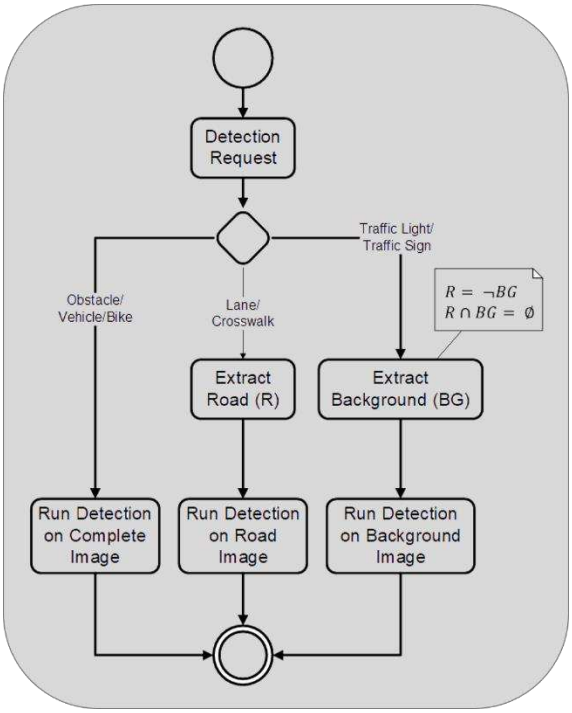

Figure 1
Partition of detections with road background segmentation

2.1 Camera-based Assistance for the Visually Impaired

We describe the state of the art of camera-based assistance for the visually impaired in [17]. The developed systems cover a wide range: In the case of sufficient remaining eyesight, camera images are represented in a suitable way (e.g. [32]). While other systems focus on the interpretation of image content in certain situations (e.g. [29]), Sensory Substitution Devices (SSD) offer a comprehensive perception of the environment by transferring the sense of sight to a different sense, e.g. hearing [7] or tactile sense [1].

In general, ASVI consist of a capture, an image processing, and an output unit [17]. The research presented in this paper focuses on the processing step and does at this stage not take the output - the way of communicating the results with the user - into account.

2.2 Road Background Segmentation

In the 1990s, road segmentation for driver assistance using morphological watersheds was introduced [5]. The algorithm uses markers situated in front of the

car; hence, it cannot be used without adaptations on data from pedestrian perspective.

From about 2008 onward, Alvarez et al. proposed road segmentation based on illumination-invariant models [2, 3, 4]. Their procedure poses two problems when used from pedestrian perspective: They use seeds in front of the car and a camera calibration is needed. The latter makes it difficult to use on any smartphone which is the platform an assistive system for the visually impaired is most likely to use.

Other algorithms are based on neural networks (e.g. [12]), support vector machines (e.g. [31]), or deep learning (e.g. [25]). These machine learning approaches are too complex and costly to be the first step in the image processing chain of an assistive system for the visually impaired.

Therefore, a road segmentation from pedestrian perspective based on Beucher et al.'s work using watersheds [5] is introduced in this article.

2.3 Crosswalk Detection

The adapted algorithm presented in the following chapter is based on the work of Choi et al. [9]. They introduce a combined detection of crosswalks and traffic lights. Their crosswalk detection is based on a 1-D mean filter in horizontal direction. Another example of an algorithm using the fact that crosswalks are horizontal structures from driver's perspective is described in [15]. Haseloff and Kummert apply Fourier and Hough Transform and thus make use of the bipolarity and straight lines which characterize crosswalks. Zhai et al. propose a crosswalk detection based on MSER and ERANSAC [37].

Furthermore, there is some research concerning crosswalk detection in the area of ASVI. The algorithms described in [28] and [33] are based on parallel lines that are extracted by Hough Transform. Cheng et al. [8] extract the bright crosswalk stripes by adaptive thresholding. They address challenging scenarios, such as partial occlusion, low contrast and distant crosswalks, and different illuminations. In addition, they offer an extensive literature review on crosswalk detection algorithms.

2.4 Lane Detection

In ASVI, the goal of lane detection is to help with general orientation by giving information about the course of the road, meaning indicating if the road goes on straight ahead or takes a left respectively right turn. On the other hand, in ADAS, lane detection is used to warn the driver if they risk to depart the lane. These lane departure warning systems consist of three steps, namely lane detection, lane tracking, and communication with the driver [24].

For ASVI, only the first step is of importance. According to [24], single frame lane detection is generally composed of four steps: (1) Image acquisition and preprocessing, (2) edge detection, (3) stripe identification by Hough Transform or Edge Distribution Function (EDF), (4) line fitting.

As the Hough Transform is not suitable to detect even slight curves, we concentrate our research on using the EDF which is the histogram of the gradient magnitude with respect to the corresponding edge angle. The work described in this paper is based on the lane detection using EDF presented by Lee [23]. After computing the region of interest with the help of the vanishing point, edge extraction as well as EDF construction are carried out. Based on an EDF analysis, Lee identifies if the car is safely within the lane or in risk of departing the lane boundaries.

3 Proposed ASVI Solutions

In the following, we describe our proposed solutions for road background segmentation as well as crosswalk and lane detection. Each reported procedure is based on known solutions from ADAS. In the end, we present an evaluation of all three procedures on a subset of the CoPeD data set [21].

In previous publications, we focused on the differences between the underlying ADAS algorithms and the proposed adaptations. This article, however, concentrates on the mathematical details of the proposed adaptations.

Table 1 shows the input and output variables for each procedure. It can be seen that road background segmentation is the foundation for the other two algorithms as the output from road background segmentation is used as input for crosswalk and lane detection.

Table 1
Input and output variables for road background segmentation, crosswalk detection, and lane detection

	Road Background Segmentation	**Crosswalk Detection**	**Lane Detection**
Input	*I*: RGB Input Image.	*I*, *R*	*I*, *R*
Output	*R*: Binary image; road pixels are white. *BG*: Binary image; background pixels are white ($BG = \neg R$).	*CW*: Binary image; crosswalk pixels are white.	*Text*: Indicates if the road is straight or takes a right/left turn, provided any lanes are detected.

In the complete chapter, (x, y) refers to a pixel position in a given image of size $N \times M$.

3.1 Road Background Segmentation

The presented algorithm is based on Beucher et al.'s work [5] and consists of five steps: After the watershed computation (1), properties of the according catchment basins are determined (2). With the help of this information, the mosaic image is built by assigning to each pixel the mean gray value of the catchment basin it belongs to (3). Afterwards, we merge adjacent catchment basins if their gradient is below a threshold (4). Finally, a decision is made which uniform region from the merged mosaic corresponds to the road part of the image and a morphological post-processing of this region is carried out (5).

Contrary to [5], we do not use a morphological gradient image as input of the watershed computation in order to reduce the number of catchment basins. To achieve this goal, we set some catchment basins to zero, depending on their mean value and mean saturation. Additionally, Beucher et al. use seeds in front of the car to expand the road from which is not possible in ASVI. To replace this procedure, we developed step (5), namely decision and morphological postprocessing.

(1) Watershed Computation

Watersheds are based on a topological interpretation of gray value images where the pixels, as spatial coordinates, are plotted against their intensity values. Local maxima are defined as watershed lines that separate the image into different regions. The inner parts of these regions are called catchment basins. In our implementation, we use the according method from Matlab's Image Processing and Computer Vision Toolbox.

In the following, we consider the notations:

I: RGB input image, smoothed with a Gaussian Filter.

gray: Gray value version of *I*.

s: s-channel (saturation) from *I* converted into HSV space.

L: Watershed image of *gray*. Every pixel has an assigned number *k* according to the catchment basin it belongs to. *L* is set to zero for watershed lines.

N: Number of catchment basins.

(2) Properties of the Catchment Basins

For every catchment basin $k, 1 \leq \mathrm{k} \leq \mathrm{N}$, we compute the mean intensity and mean saturation value:

$$\overline{gray}(k) = \sum_{\{(x,y)|L(x,y)=k\}} \frac{gray(x,y)}{\#\{(x,y)|L(x,y) = k\}}$$

$$\bar{s}(k) = \sum_{\{(x,y)|L(x,y)=k\}} \frac{s(x,y)}{\#\{(x,y)|L(x,y) = k\}}$$

Thereby, # stands for the number of elements in a set.

(3) Mosaic Image

For $L(x,y) = k$, the mosaic image is then defined as

$$mosaic(x,y) = \begin{cases} \overline{gray}(k), \text{if } k \neq 0 \wedge \bar{s}(k) < th_s \wedge th_1 < \overline{gray}(k) < th_2 \\ 0, \text{ else} \end{cases}.$$

Depending on the thresholds th_s, th_1, and th_2, the value of the mosaic image is either set to the mean gray value or to zero.

If $mosaic(x,y)$ is set to zero, we simultaneously set $L(x,y)$ to zero. This means that the value of each pixel is set to the mean gray value of the corresponding catchment basin, provided mean gray value and mean saturation of the basin are in ranges that make them possible parts of the road.

(4) Merged Mosaic Image

In this step, adjacent catchment basins are merged, if the gradient between them falls below a threshold th_{gray}. Let k and l be the identification numbers of adjacent catchment basins. Their gradient is then defined as:

$$\nabla(k,l) = |\ \overline{gray}(k) - \overline{gray}(l)|$$

If it holds that $\nabla(k,l) < th_{gray}$, we update $\overline{gray}(k)$ and $\overline{gray}(l)$ to their mean value

$$\overline{gray}(k) = \overline{gray}(l) = \frac{\overline{gray}(k) + \overline{gray}(l)}{2}$$

and the mosaic image to the merged mosaic image

$$mosaic(x,y) = \begin{cases} \overline{gray}(k), \text{if } (L(x,y) = k \vee L(x,y) = l) \wedge \nabla(k,l) < th_{gray} \\ mosaic(x,y), \text{ else} \end{cases}.$$

Furthermore, we set $L(x,y) = k$ for all $\{(x,y)|L(x,y) = l\}$.

We repeat this process for all adjacent catchment basins.

(5) Decision and Morphological Postprocessing

Let max_i be the gray value that occurs most often in the merged mosaic image. We then define the binary road image *R* as:

$$R(x,y) = \begin{cases} 1, \text{if } mosaic(x,y) = max_i \\ 0, \text{ else} \end{cases}$$

We perform morphological postprocessing by applying filling, opening, largest component, and closing on R.

The background is then determined as the complement of R:

$$BR = \neg R$$

For stability reasons, we use the union of the last n frames as a tracking step. Hence, the road of the n-th frame in a sequence is defined as:

$$R_n = \bigcup_{i=0}^{n-1} R_{n-i}$$

In tests, $n = 10$ led to good results.

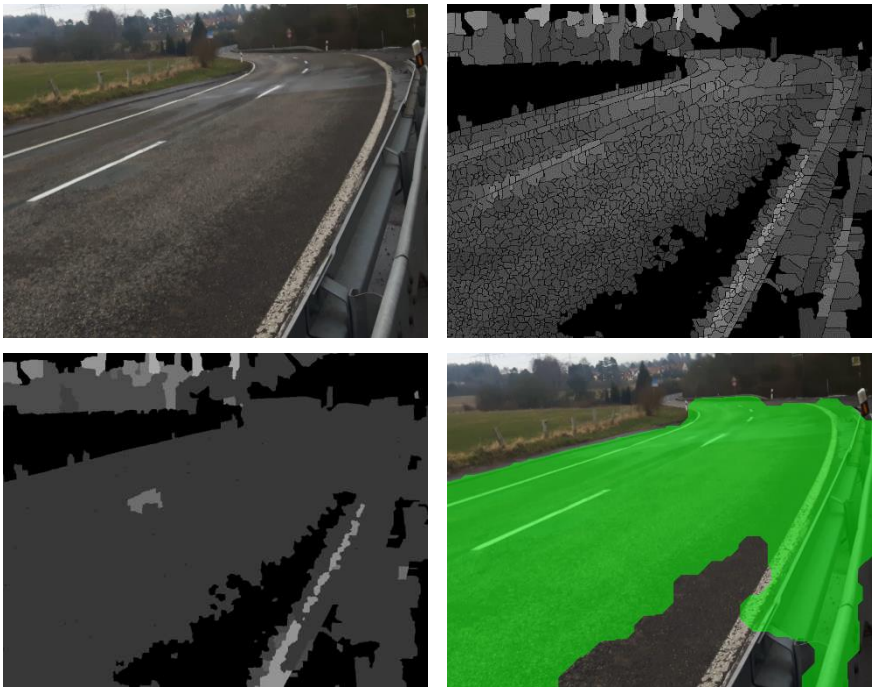

Figure 2
Intermediate steps of road background segmentation (cropped images):
Original, mosaic, merged mosaic, and result

Figure 2 shows intermediate steps of this algorithm for an example image. Even though, the road is not completely extracted, the segmentation counts as successful, because there are enough road detail to determine the road path. Furthermore, it is important that the bus stop sign (right side) was not extracted as road.

3.2 Crosswalk Detection

The presented algorithm is based on Choi et al.'s work [9] and preliminary descriptions can be found in [19] and [21]. It consists of four steps: After preprocessing and computing the Region of Interest (ROI) (1), horizontal and vertical 1-D mean filters are applied (2). The differences between original and filtered images are the foundation for binarization and morphological postprocessing (3). Finally, the two resulting masks are combined via a bitwise or operation and a decision is made if the remaining pixels form a crosswalk or not (4). According to Table 1, the original image as well as the extracted road from the road segmentation algorithm are the input variables for this algorithm.

Choi et al. [9] developed their algorithm for an ADAS containing a tilted front camera so that the resulting images show only the road, but no background. Therefore, no ROI was needed in [9] and we developed the before presented road background segmentation as ROI. Furthermore, we added a vertical filter in order to be able to detect crosswalks from every angle by combining it with the horizontal filter. As we consider two filters, our handling of the masks differs from the one in [9]. Finally, Choi et al. [9] gave no detailed description of their postprocessing and decision step so that we had to develop our own version.

(1) Preprocessing and ROI

As preprocessing, we apply a Gaussian filter for noise reduction. Alternatively, we use the already smoothed RGB image I from the road segmentation algorithm as input.

Let $gray$ be the gray value version of I. The ROI is defined by the before computed road R. We therefore set the background pixels in $gray$ to zero and keep the values belonging to the road:

$$gray(x,y) = \begin{cases} gray(x,y), \text{if } R(x,y) = 1 \\ 0, \text{else} \end{cases}$$

(2) 1-D Mean Filters

Choi et al. [15] propose a horizontal 1-D mean filter to detect crosswalks, because from driver's perspective crosswalks are horizontal structures. For pedestrians, however, crosswalks appear in every possible angle. Therefore, we use an additional 1-D mean filter in vertical direction. By combining the two filters, we are able to detect crosswalks occurring in any angle.

In order to respect the ROI, pixels having the value zero are excluded from the computation. With that, we get the result for filtering in horizontal respectively vertical direction as G_x and G_y:

$$G_x(x,y) = \sum_{k=-s}^{s} \frac{gray(x+k,y)}{\#\{(x+k,y)| -s \le k \le s \wedge gray(x+k,y) \neq 0\}},$$

$$G_y(x,y) = \sum_{l=-s}^{s} \frac{gray(x,y+l)}{\#\{(x,y+l)| -s \le l \le s \wedge gray(x,y+l) \neq 0\}}.$$

The size of the filter is in both cases $(2 \cdot s + 1)$.

(3) Binarization and Morphological Postprocessing

We compute the differences between the original image $gray$ and the filtered results G_x and G_y as D_x and D_y:

$$D_x(x,y) = |gray(x,y) - G_x(x,y)|$$

$$D_y(x,y) = \left|gray(x,y) - G_y(x,y)\right|$$

By applying a threshold th, we get the masks M_x and M_y:

$$M_x(x,y) = \begin{cases} 1, \text{if } D_x(x,y) > th \\ 0, \text{else} \end{cases}$$

$$M_y(x,y) = \begin{cases} 1, \text{if } D_y(x,y) > th \\ 0, \text{else} \end{cases}$$

Afterwards, we perform closing followed by opening with a larger structuring element on both masks independently.

(4) Bitwise or Operation and Decision

Before combining the two masks, we delete components that cover less than 1.5 % of the image. We then unite the two masks to one mask M by applying a pixelwise or operation:

$$M(x,y) = M_x(x,y) \vee M_y(x,y)$$

The largest component CW of M is a crosswalk candidate, provided it also exceeds the 1.5 % mark. The decision is made based on two values: The extent of CW indicating the relative number of pixels that are set inside a box surrounding CW and the ratio between minor and major axis length of CW.

If both, ratio and extent, are higher than thresholds th_r and th_e, CW is considered a crosswalk. Otherwise, no crosswalk is present and CW is set to zero.

Figure 3 shows intermediate results of the crosswalk detection algorithm for an example image. The figure only shows intermediate steps for vertical direction, because the crosswalk in this example is a vertical structure. Therefore, the horizontal mask is zero and the vertical mask is identical with the result.

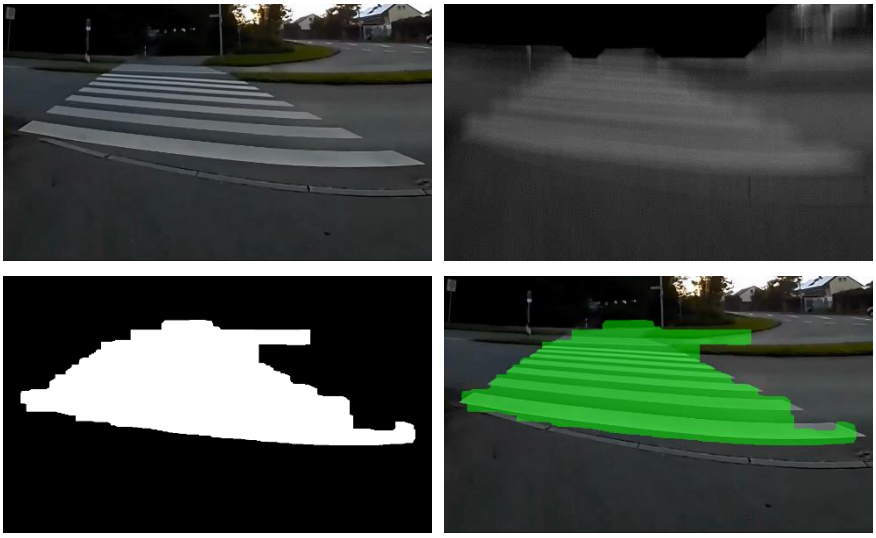

Figure 3
Intermediate steps of the presented crosswalk detection (cropped images):
Original, vertical filter, vertical mask, and result

3.3 Lane Detection

The presented algorithm is based on Lee's work [23] and a short description was published in [20]. Six subsequent steps are carried out: After preprocessing and computing the ROI based on the road background segmentation (1), we divide the ROI into a total of eight subimages (2). For every subimage, we compute (3) and analyze (4) the EDF, resulting in an angle for every subimage. Afterwards, the angles are interpolated (5) and the course of the road ahead is decided (6) based on the concavity of the interpolated function.

Lee's ROI computation [23] is based on the vanishing point. Because we cannot use this procedure in ASVI, we use the before extracted road as ROI. In contrast, EDF computation and analysis were mostly taken from [23]. The partition into subimages and the decision step were newly developed by us. As the purposes of lane detection differ for ADAS and ASVI, there are no according steps in [23]. Interpolation by linear parabolic fitting is carried out similarly to [24], but was adapted to the presented procedure.

(1) Preprocessing and ROI

This step is analogous to the before presented crosswalk detection algorithm. Additionally, the first row of the road R containing a non-zero entry is computed as min_y.

(2) Partition into Subimages

We divide the image from the bottom of the image to min_y into eight parts. From bottom to top, the first two subimages consist of one fourth of the available rows, the next two of one eighth, and the last four of one sixteenth.

The following two steps are carried out for each of the eight subimages. This will be indicated with the index j, where $j = 1$ refers to the subimage at the bottom which is the one closest to the user.

(3) EDF Computation

We compute the gradient $G_j(x,y) = \left(G_{j_x}(x,y), G_{j_y}(x,y)\right)^T, j = 1, \dots, 8,$ by convolving the image with the Sobel masks:

$$H_x = \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix}$$

$$H_y = H_x^T$$

The magnitude $M_j(x,y)$ is then defined as

$$M_j(x,y) = \sqrt{G_{j_x}^2(x,y) + G_{j_y}^2(x,y)}, j = 1, \dots, 8$$

and the angle $\alpha(x,y)$ is computed as

$$\alpha(x,y) = \tan^{-1}\left(\frac{G_{j_y}(x,y)}{G_{j_x}(x,y)}\right), j = 1, \dots, 8,$$

so that the result is in the range $[1,180]$. The degree values are rounded to integers.

Before determining the EDF, we compute the 97%-quantile of M and set the values below it to zero. The EDF is then definied as:

$$EDF_j(i) = \sum_{i=\alpha(x,y)} M(x,y), i = 1, \dots, 180\,, j = 1, \dots 8$$

(4) EDF Analysis

First, we smooth $EDF_j, j = 1, \dots, 8,$ with a 1D-filter of size 15. Afterwards, we compute the relative value of the highest peak $p_j \in [0,1], j = 1, .., 8$ and its position $\theta_j \epsilon [1,180], j = 1, \dots, 8$. In case of no occurring peak, we set $p_j = 0$ and $\theta_j = 180$.

(5) Interpolation: Linear Parabolic Fitting

For interpolation, we consider the lane in the first two subimges from the bottom, the ones closest to the user, as linear and the remaining lane as parabolic. In order to be able to interpolate independently from image size, we normalize the values to $[0,1]$. Furthermore, we change the axes, meaning that the x-values are defined according to the size computed in step (2). With that, we get the interpolation function

$$f(x) = \begin{cases} a \cdot (x - 0.5) + b, \text{if } x \leq 0.5 \\ a \cdot (x - 0.5) + b + c \cdot (x - 0.5)^2, \text{else} \end{cases}$$

and the x-values $x = (0, 0.25, 0.5, 0.625, 0.75, 0.8125, 0.875, 0.9375, 1)^T$.

The according y-values are computed by determining a line with angle θ_j:

$$y_1 = 0$$

$$y_j = y_{j-1} + (x_j - x_{j-1}) \cdot \tan(\theta_{j-1} - 90), j = 2, \dots, 9$$

It is necessary to subtract 90 from the angle, because we rotated the axes. After computing, the y-values, we normalize them to $[0,1]$.

To get the interpolation function f, we compute the mean square error solution of the following overdetermined system of equations:

$$\begin{pmatrix} x_1 - 0.5 & 1 & 0 \\ x_2 - 0.5 & 1 & 0 \\ x_3 - 0.5 & 1 & 0 \\ x_4 - 0.5 & 1 & (x_4 - 0.5)^2 \\ x_5 - 0.5 & 1 & (x_5 - 0.5)^2 \\ x_6 - 0.5 & 1 & (x_6 - 0.5)^2 \\ x_7 - 0.5 & 1 & (x_7 - 0.5)^2 \\ x_8 - 0.5 & 1 & (x_8 - 0.5)^2 \\ x_9 - 0.5 & 1 & (x_9 - 0.5)^2 \end{pmatrix} \cdot \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{pmatrix}$$

(6) Decision

There are four possible outcomes for the algorithm: The input image does not contain lanes, the road goes straight ahead, or it takes a right or left turn.

First, we determine if the image contains lanes by checking if the mean value of the highest peaks $\bar{p} = \frac{1}{8}\sum_{j=1}^{8} p_j$ is within a certain range $[th_{p_{low}}, th_{p_{high}}]$.

This is based on the idea that there is a characteristic amount of pixels belonging to the lane boundaries that indicate the direction. In tests, these pixels made up between 1.4% and 3% of the subimage.

If $\bar{p}$ lies outside the specified range, the output text is set to 'No lane detected'. Otherwise, we check the course of the road by examining the parameter c from the interpolation function and the variance of the θ-values $\bar{\bar{\theta}} = \frac{1}{7} \cdot \sum_{j=1}^{8} (\theta_j - \bar{\theta})$, where $\bar{\theta}$ is the mean θ value. In general, a negative c value means that the parabola is concave down and the road takes a right turn; accordingly, a positive value of c means that the parabola is concave up and the road takes a left turn. The higher the absolute value of c, the tighter the curve of the road; the lower the absolute value of c, the straighter the road. At the same time, the variance of the angles within the image is higher when the curve is tighter. With this knowledge, we set up a four-step process to determine the course of the road depending on several thresholds:

(1) If $|c| < th_{c_{low}}$, the output text is 'Straight ahead'.

(2) If $th_{c_{low}} \leq |c| < th_{c_{mid}}$, the output is determined as:

$$\text{text} = \begin{cases} \text{'Straight ahead'}, \text{if } \bar{\bar{\theta}} < th_{var_{high}} \\ \text{'Left turn'}, \text{if } \bar{\bar{\theta}} \geq th_{var_{high}} \wedge c > 0 \\ \text{'Right turn'}, \text{else} \end{cases}$$

(3) If $th_{c_{mid}} \leq |c| < th_{c_{high}}$, the output is determined as:

$$\text{text} = \begin{cases} \text{'Straight ahead'}, \text{if } \bar{\bar{\theta}} < th_{var_{low}} \\ \text{'Left turn'}, \text{if } \bar{\bar{\theta}} \geq th_{var_{low}} \wedge c > 0 \\ \text{'Right turn'}, \text{else} \end{cases}$$

(4) If $|c| \geq th_{c_{high}}$, the output is determined as:

$$\text{text} = \begin{cases} \text{'Left turn'}, \text{if } c > 0 \\ \text{'Right turn'}, \text{else} \end{cases}$$

To increase the robustness, we compute the mean values of $\bar{p}$ and $\bar{\bar{\theta}}$ for the last 15 frames and set the output text to the one occurring the most in the last 15 frames.

Figure 4 shows the interpolated function for an example image. In general, the interpolated function does not match the course of the road exactly, but is a good approximation and has identical concavity.

3.4 Evaluation

The three before described algorithms were implemented in *Matlab Version R2017b* using the *Image Processing and Computer Vision Toolbox*. The output of the road segmentation was used as input for the ROI computation of crosswalk and lane detection.

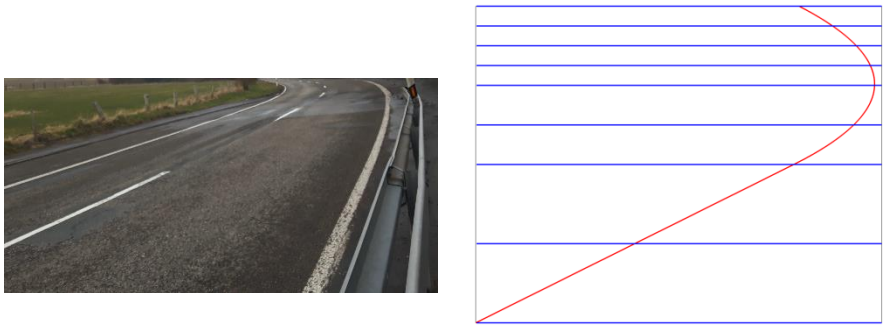

Figure 4
Lane detection: Original (cropped image) and interpolated function.
Blue lines mark the subimages.

The algorithms were tested on a subset of the CoPeD data set [21]. In the following, we give the relative path to the used sequences inside the folder that can be downloaded from the CoPeD website (Link: see Chapter 1):

- Crosswalk with traffic (CW w/ traffic), first 243 frames:
  CoPeD\2 Crossings\Pedestrian\ Pedestrian_Crosswalk_Traffic.mp4
- Crosswalk without traffic (CW w/o traffic), first 195 frames:
  CoPeD\2 Crossings\Pedestrian\ Pedestrian_Crosswalk_NoTraffic.mp4
- Straight, 74 frames:
  CoPeD\1 Lane Detection\Pedestrian\Others\ Pedestrian_Straight_2.mp4
- Left, 489 frames:
  CoPeD\1 Lane Detection\Pedestrian\Others\ Pedestrian_Left_2.mp4
- Right, 585 frames:
  CoPeD\1 Lane Detection\Pedestrian\Others\ Pedestrian_Right_2.mp4

Tables 2, 3, and 4 show the evaluation for road segmentation (Table 2), crosswalk detection (Table 3), and lane detection (Table 4). In the first two cases, we considered nine frames less than stated above, in the third case 14 frames less. This is because we computed the union of the last ten frames for road segmentation and formed the mean of different values of the last 15 frames for lane detection.

For road segmentation, an over detection means that in addition to the road, important data belonging to the background according to Figure 1 (traffic lights and signs) close to the user, at least 20×20 pixels in size, were extracted. An under detection means that not enough of the road was extracted in order to detect the road's markings.

For crosswalk detection, the goal was to detect the crosswalk in all frames of the first two sequences and no crosswalk in the other sequences. In the case of lane detection, it was the other way around: For the first two sequences, the result

should be 'No lane detected' whereas for the latter three the respective course of the road should be determined.

Table 2
Evaluation Road Background Segmentation (NF: Number of Frames, CD: Correct Detection, OD: Over Detection, UD: Under Detection)

	CW w/ traffic	CW w/o traffic	Straight	Left	Right	**Total**
NF	234	186	65	480	576	1541
CD	234	186	65	478	576	1539
OD	0	0	0	2	0	2
UD	0	0	0	0	0	0
Hit rate	**100.00%**	**100.00%**	**100.00%**	**99.59%**	**100.00%**	**99.87%**

Table 3
Evaluation Crosswalk Detection (NF: Number of Frames, CD: Correct Detection, ND: Not Detected if crosswalk is present, FP: False Positive)

	CW w/ traffic	CW w/o traffic	Straight	Left	Right	**Total**
NF	234	186	65	480	576	1541
CD	226	181	65	472	576	1520
ND	8	5	0	0	0	13
FP	0	0	0	8	0	8
Hit rate	**96.71%**	**97.37%**	**100.00%**	**98.36%**	**100.00%**	**98.64%**

Table 4
Evaluation Lane Detection (NF: Number of Frames, CD: Correct Detection, WD: Wrong Detection)

	CW w/ traffic	CW w/o traffic	Straight	Left	Right	**Total**
NF	229	181	60	475	571	1516
CD	229	181	60	473	541	1484
WD	0	0	0	2	30	32
Hit rate	**100.00%**	**100.00%**	**100.00%**	**99.58%**	**94.75%**	**97.89%**

The overall hit rates were 99.87% for road segmentation, 98.64% for crosswalk, and 97.89% for lane detection. It would be useful to compare the evaluations with results from the underlying ADAS algorithms, but this is difficult to achieve because they were tested on other, not publicly available data. Choi et al. [9] report 96.2% correctly detected, present crosswalks and 0.66% false positives for a data set of 21864 frames from which 1053 contain a crosswalk. In our case, the according numbers were with 96.9% and 0.76% in a similar range. For lane detection, Lee [23] tested on 1200 frames, if the car departed from the road or not. The reached hit rate was 96.42%. Beucher et al. [5] presented their results on a

small number of single frames. Even though it is not possible in all cases to compare our hit rates with the ones of the underlying ADAS algorithms, we can state that our adapted algorithms performed well and reached suitable hit rates that are applicable for ASVI.

4 Future Work

In the future, we will implement the ADAS algorithms on which the presented algorithms are based in order to be able to compare performance in the fields of ADAS and ASVI. The evaluation will be done using the CoPeD data set [21], which contains comparable video sequences from the driver and pedestrian perspectives. Implementing the ADAS algorithms is a challenging task, because the algorithms are in general not described in detail in the related papers.

Another possibility to provide information about the course of the road to a visually impaired pedestrian, is to adapt the ADAS algorithm from [11] which detects arrows on the road setting out lane arrangements.

Improving the segmentation will also improve the crosswalk and lane detection results. Therefore, we will consider up-to-date road segmentation algorithms using deep learning (e.g. [25]).

Additionally, we will focus on the remaining paths in Figure 1 and work on ASVI solutions based on ADAS algorithms for obstacle, especially bike and vehicle, detection (e. g. [13], [22], [35]) as well as traffic light (e. g. [27]) and traffic sign detection (e. g. [36]). The adaptations for all considered use cases will be summarized in a concept for transferring ADAS algorithms to ASVI using methods from software engineering [30] and project management [26].

Conclusions

The content in this article is embedded in our project of developing a concept for the transfer of camera-based algorithms from ADAS to ASVI.

In previous work, we identified a total of seven use cases that have to be considered. This article focuses on the detection of markings on the road, namely crosswalk and lane detection. In order to reduce the computational cost, we first introduced a road background segmentation. Thereby, crosswalk and lane detection can be carried out on the road only. Algorithms for traffic sign and traffic light detection, to be developed in the future, will be applied to the background part of the images, whereas obstacles will be detected in the whole image.

The presented road background segmentation is based on Beucher et al.'s [4] use of morphological watersheds and in addition used thresholds for the mean gray

and saturation values of each catchment basin. Our model reached a hit rate of 99.87%.

The basis of our developed crosswalk detection is Choi et al.'s idea of using a horizontal mean filter [7]. We combined the horizontal with a vertical filter and defined a decision process. The recognition rate of this algorithm was 98.64%.

The suggested lane detection used the Edge Distribution Function (EDF). Instead of applying the EDF to the whole image as in Lee's ADAS work [18], we first divided the image into subimages of increasing sizes from bottom to top. Afterwards, interpolation of the angles and analysis of the according function returned the course of the road. Correct detections occurred in 97.89% of the examined frames.

With that, we can state that the presented ASVI algorithms, adapted from relevant ADAS methods, achieved overall good hit rates and thus are applicable for ASVI.

In our future work, we will first compare the performance of the presented ASVI algorithms with their underlying ADAS solutions and then focus on the remaining use cases.

References

[1] M. R. Adame, J. Yu, K. Moeller: Mobility support system for elderly blind people with a smart walker and a tactile map. Proceedings of XIV Mediterranean Conference on Medical and Biological Engineering and Computing, pp. 608-613, April, 2016

[2] J. M. Alvarez, A. Lopez, R. Baldrich: Illuminant-invariant model-based road segmentation. Proceedings of the IEEE Intelligent Vehicles Symposium, pp. 1175-1180, June, 2008

[3] J. M. Alvarez, A. M. Lopez: Road detection based on illuminant invariance. IEEE Transactions on Intelligent Transportation Systems, Vol. 12, No. 1, pp. 184-193, March, 2011

[4] J. M. Alvarez, T. Gevers, Y. LeCun, A. M. Lopez: Road scene segmentation from a single image. Proceedings of the 12th European Conference on Computer Vision, pp. 376-389, Springer, October, 2012

[5] S. Beucher, M. Bilodeau, X. Yu: Road segmentation by watershed algorithms. Proceedings of the Pro-art vision group PROMETHEUS workshop, pp. 1-24, April, 1990

[6] R. R. Bourne, S. R. Flaxman, T. Braithwaite, M. V. Cicinelli, A. Das, J. B. Jonas et al.: Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. The Lancet Global Health, Vol. 5, No. 9, pp. 888-897, September, 2017

[7] S. Caraiman, A. Morar, M. Owczarek, A. Burlacu, D. Rzeszotarski, N. Botezatu, P. Herghelegiu, F. Moldoveanu, P. Strumillo, A. Moldoveanu: Computer vision for the visually impaired: the sound of vision system. Proceedings of the IEEE International Conference on Computer Vision Workshops, pp. 1480-1489, October, 2017

[8] R. Cheng, K. Wang, K. Yang, N. Long, W. Hu, H. Chen, J. Bai, D. Liu: Crosswalk navigation for people with visual impairments on a wearable device. Journal of Electronic Imaging, Vol. 26, No. 5, pp. 053025-1 – 053025-14, September/October, 2017

[9] J. Choi, B. T. Ahn, I. S. Kweon: Crosswalk and traffic light detection via integral framework. Proceedings of the 19th IEEE Korea-Japan Joint Workshop on Frontiers of Computer Vision, pp. 309-312, January/ February, 2013

[10] P. S. Duckett, R. Pratt: The researched opinions on research: visually impaired people and visual impairment research. Disability & Society, Vol. 16, No. 6, pp. 815-835, October, 2001

[11] Z. Fazekas, P. Gáspár: Computerized recognition of traffic signs setting out lane arrangements. Acta Polytechnica Hungarica, Vol. 12, No. 5, pp. 35-50, 2015

[12] M. Foedisch, A. Takeuchi: Adaptive real-time road detection using neural networks. Proceedings of the 7th International IEEE Conference in Intelligent Transportation Systems, pp. 167-172, October, 2004

[13] D. Greene, J. Liu, J. Reich, Y. Hirokawa, A. Shinagawa, H. Ito, T. Mikami: An efficient computational architecture for a collision early-warning system for vehicles, pedestrians, and bicyclists. IEEE Transactions on Intelligent Transportation Systems, Vol. 12, No. 4, pp. 942-953, 2011

[14] J. Fritsch, T. Kuehnl, A. Geiger: A new performance measure and evaluation benchmark for road detection algorithms. Proceedings of International IEEE Conference on Intelligent Transportation Systems, pp. 1693-1700, October, 2013

[15] A. Haselhoff, A. Kummert: On visual crosswalk detection for driver assistance systems. Proceedings of the IEEE Intelligent Vehicles Symposium, pp. 883-888, June, 2010

[16] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, C. Igel: Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark. Proceedings of International Joint Conference on Neural Networks, pp. 1-8, August, 2013

[17] J. Jakob, E. Cochlovius, C. Reich: Kamerabasierte Assistenz für Blinde und Sehbehinderte. InformatikJournal 2016/17, pp. 3-10, 2016, Available: https://tinyurl.com/informatikJournal2016

[18] J. Jakob, K. Kugele, J. Tick: Defining camera-based traffic scenarios and use cases for the visually impaired by means of expert interviews. Proceedings of the IEEE 14th International Scientific Conference on Informatics, pp. 128-133, November, 2017

[19] J. Jakob, J. Tick: Concept for transfer of driver assistance algorithms for blind and visually impaired people. Proceedings of the IEEE 15th International Symposium on Applied Machine Intelligence and Informatics, pp. 241-246, January, 2017

[20] J. Jakob, J. Tick: Traffic scenarios and vision use cases for the visually impaired. Acta Electrotechnica et Informatica, Vol. 18, No. 3, pp. 27-24, September, 2018

[21] J. Jakob, J. Tick: CoPeD: Comparable Pedestrian Driver data set for traffic scenarios. Proceedings of the IEEE 18th International Symposium on Computational Intelligence and Informatics, pp. 87-92, November, 2018

[22] G. Kertész, S. Szénási, Z. Vámossy: Mulit-directional image projections with fixed resolution for object matching. Acta Polytechnica Hungarica, Vol. 15, No. 2, pp. 211-229, 2018

[23] J. W. Lee: A machine vision system for lane departure detection. Computer Vision and Image Understanding, Vol. 86, No. 1, pp. 52-78, 2002

[24] R. P. Loce, R. Bala, M. Trivedi: Computer vision and imaging in intelligent transportation systems. John Wiley & Sons Ltd, West Sussex, 2017

[25] G. L. Oliveira, W. Burgard, T. Brox: Efficient deep models for monocular road segmentation. Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp. 4885-4891, October, 2016

[26] P. Rosenberger, J. Tick: Suitability of PMBOK 6th edition for agile developed IT projects. Proceedings of the 18th International Symposium on Computational Intelligence and Informatics, pp. 241-246, November, 2018

[27] S. Saini, S. Nikhil, K. R. Konda, H. S. Bharadwaj, N. Ganeshan: An efficient vision-based traffic light detection and state recognition for autonomous vehicles. Proceedings of the IEEE Intelligent Vehicles Symposium, pp. 606-611, June, 2017

[28] S. Se: Zebra-crossing detection for the partially sighted. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2211-2217, June, 2000

[29] R. Shilkrot, J. Huber, R. Boldu, P. Maes, S. Nanayakkara: FingerReader: A finger-worn assistive augmentation. In Assistive Augmentation, pp. 151-175, Springer, Singapore, 2018

[30] I. Sommerville: Software Engineering. Pearson, Boston, 2011

[31] M. Song, D. Civco: Road extraction using SVM and image segmentation. Photogrammetric Engineering & Remote Sensing, Vol. 70, No. 12, pp. 1365-1371, December, 2004

[32] J. J. Van Rheede, I. R. Wilson, L. Di Bon-Conyers, S. Croxford, R. E. MacLaren, S. L. Hicks: Smart Specs: Electronic vision enhancement in real-life scenarios. Investigative Ophthalmology & Visual Science, Vol. 57, No. 12, p. 5165, 2016

[33] S. Wang, Y. Tiang: Detecting stairs and pedestrian crosswalks for the blind by RGBD camera. Proceedings of IEEE International Conference on Bioinformatics and Biomedicine Workshops, pp. 732-739, October, 2012

[34] A. Witzel: The problem-centered interview. Forum: Qualitative Social Research, Vol. 1, No. 1, Art. 22, January, 2000

[35] M. T. Yang, J. Y. Zheng: On-road collision warning based on multiple foe segmentation using a dashboard camera. IEEE Transactions on Vehicular Technology, Vol. 64, No. 11, pp. 4974-4984, 2015

[36] Y. Yuan, Z. Xiong, Q. Wang: An incremental framework for video-based traffic sign detection, tracking, and recognition. IEEE Transactions on Intelligent Transportation Systems, Vol. 18, No. 7, pp. 1918-1929, 2017

[37] Y. Zhai, G. Cui, Q. Gu, L. Kong: Crosswalk detection based on MSER and ERANSAC. Proceedings of the IEEE 18th International Conference on Intelligent Transportation Systems, pp. 2770-2775, September, 2015

Active-Reactive Power Control of Grid-Tied Solar Energy Systems, Using Multi-Parameter Band Control for Grid Voltage Regulation

Mustafa Tekin*, Mustafa Sekkeli

University of Kahramanmaras Sutcu Imam, Faculty of Engineering and Architecture, Department of Electrical-Electronic Engineering, Kahramanmaras, Turkey, email: mustafatekin@ksu.edu.tr, msekkeli@ksu.edu.tr

Abstract: One of the most studied and funded renewable energy sources is solar energy, all over the world. Solar panels are used to obtain electrical energy from the sun. They should be used effectively due to their expensive structure and challenging installation. However, as the sun light only reaches to the earth during daylight hours, solar systems cannot generate active power during the night, or in cases where daylight is not enough, the whole power capacity of the solar inverters have not been used. In this study; it was shown that in addition to the active power generation during the daytime hours, the solar systems were able to generate reactive power with the remaining capacity of the inverter. Thereby adjusting the grid voltage was provided by the solar system's reactive power control. Designed solar system generates completely reactive power at nights. Thus, the efficiency of the solar systems has been increased by ensuring that the solar systems have used to keep the grid voltage within the values, which stated in the standards, by producing reactive power as well as active power. Matlab/Simulink was used to model the system when using hysteresis band method for system control.

Keywords: power grids; reactive power control; renewable energy sources; solar system; voltage control

1 Introduction

Daily, the development of technology and the increase in the human population increases the demand for electrical energy. People have used traditional sources of energy, such as, coal, natural gas and oil for years, to meet electricity demands. Today, however, these resources are not capable in meeting the energy needs, moreover, using these resources pollutes the environment and depletes its reserves. Therefore, over the last twenty years, renewable energy sources which are solar, wind, wave, geothermal and biogas have become very popular and investments have focused on these resources. One of the most popular and most

important among these renewable sources is solar energy, because of its simple and enduring structure. Photons from the sun have been converted into electrical energy through photovoltaic panels. The fact that used photovoltaic panels and the inverters are not cheap enough yet, emphasize the necessity of using these systems as efficiently as possible. The biggest disadvantages of using photovoltaic panels are that they do not generate at night due to the lack of sunlight on the earth and it generates an unstable amount of energy differ with the changing sunlight conditions during the day. These disadvantages can be eliminated by ensuring that the inverters used in solar systems can regulate voltage by producing reactive power at night with all their capacities and they can generate reactive power along with active power during the daytime, through the inverter's remaining capacity. Scientists have been researching the possibilities of using the full capacity of the inverters used in systems. Because they are completely non-functional at night and not employed at full capacity during the day. Some relavent current studies are listed below.

Q/V control was performed to prevent the voltage oscillations caused by the photovoltaic systems which connected to the grid in [1]. In [2], the authors developed a rapid power flow response by developing an algorithm that generates a reference current. In [3], The voltage control of the busbar at the end of the line, which was selected as the critical busbar in the grid where many PhotoVoltaic (PV) systems were connected, was made by using reactive power and necessary control technique. It was performed reactive power control of a single phase, transformerless, grid connected solar inverter by [4]. In [5], voltage fluctuations in the grid due to solar systems were discussed and this problem was solved by reactive power control. In [6], the unnecessary switching of the On-Line Tap Changer (OLTC) was prevented by ensuring that the OLTC, which was present in the grid and PV system were working compatibly with each other. Using the sliding mode control, the inverter in the PV system was provided to generate useful reactive power in the study [7]. In article [8], the reactive power at the inverter output was controlled by the linear decision rules method. In [9] and [10], the active and reactive power control of the inverter was realized. In [11], to prevent the voltage rise at the grid connection point of the PV systems, the reactive power adjusted through the inverter's reactive current control. In article [12], the reactive power of the inverter was controlled by the Quasi Sinusoidal Waveform method. The triggering performances of the solar inverter's, used as reactive power generator, switching elements were analyzed in [13]. After the connection of the PV systems to the grid, the reverse power flow due to the supply-demand imbalance was mentioned and possible solution for the situation can be fixed with the control of *P-Q* and *Q-V* in [14]. In [15], the solar inverter was used as a harmonic and reactive power compensator. The active-reactive power production balance of the inverter output was controlled in [16-18]. The reactive power control of the inverter was made by reactive current control in [19]. In [20], the response of the inverter to the unexpected conditions in the grid was analyzed. In the study [21], the PV system with a battery was connected in

parallel with the grid and the inverter of PV provided harmonic compensation to the grid using the active-reactive power control. In [22], active harmonic filtering was performed by using type-2 fuzzy control method. The active-reactive power control of the inverter was done by the passivity-based method in the article [23], it was made by the incremental conductance method in [24]. A new relay was designed for reactive power control in [25].

In the light of all these works; the grid-connected PV system was modelled in Matlab/Simulink environment in this study. By controlling the phase angle and amplitude of the inverter voltage of this PV system, in case of the presence of the insufficient amount of sunlight or at night, it was ensured that the grid voltage was kept at optimum values with the active-reactive power production balance of the inverter. Unlike other studies in the literature, with this work, it was ensured that the inverter provides the most suitable active-reactive power response to the grid by considering the amount of radiation on the environment, the inverter power capacity and whether the grid voltage was at the critical limit values. Accordingly, a new approach to the literature has been introduced.

2 Structure and Components of the System

In this study; a low-voltage grid, which is connected to the three-phase solar system, was modelled. The capacity of the inverter connected to the electricity grid was selected as 1 kVA in the Matlab model. The PV system consists of a phase lock loop (PLL), a pulse width modulation (PWM) generator, solar inverter and controller of this inverter. The system diagram was given in Figure 1.

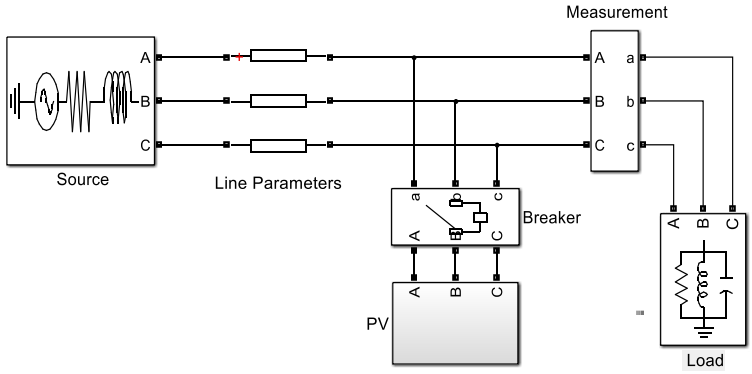

Figure 1
Power system with solar inverter

2.1 Solar System

The model of the PV system was given in Figure 2. The PLL and PWM generator of the inverter were embedded in the PWM generator block. A boost type converter was used in the solar system and the inverter output was filtered as it was shown in Figure 2. Phase angle control was performed with phase controller block. There is a system in the grid and it determines the amount of radiation, the inverter capacity and the reactive power in the phase controller block.

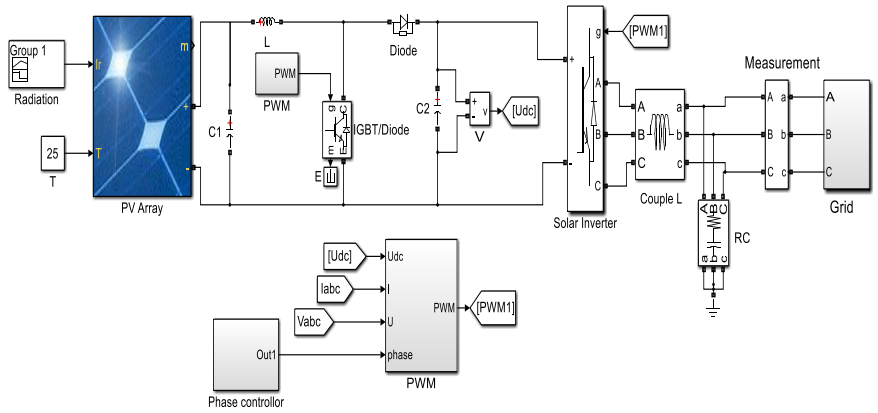

Figure 2
Solar system

2.2 Phase Lock Loop (PLL)

In the grid-connected solar system, PLL is the component that allows the inverter and grid voltage to synchronize with each other [26]. A PLL operates in its simplest form that the phase angle of the reference signal from the grid is determined by the phase detector. Then the obtained signal is passed through the low pass filter and entered into the voltage-controlled oscillator so that the voltage is generated at the desired phase angle and frequency. The PLL block was shown in Figure 3.

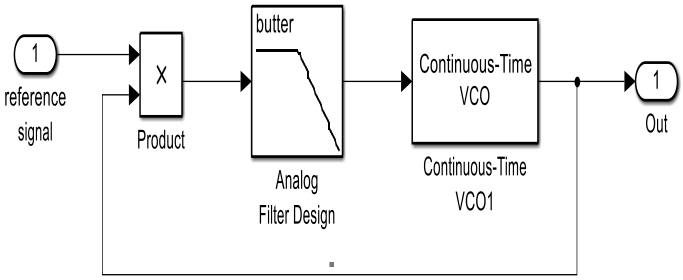

Figure 3
PLL block diagram

2.3 PWM Generator

The PWM is a control element that converts analogue input to digital output (0 or 1). PWM technique has been used in many areas. One of the most important parameters in this technique is Duty Cycle (D) [27]. D is a variable that analyzes how long the PWM output is "on" over a period and is usually given in percentage.

The logic operation of PWM was summarized in Figure 4. According to this figure; the triangular wave has compared to the sinus wave, during the time when the sinus is greater than the triangle, the output is equal to 1, during the time when vice versa output is equal to 0 and thus sine wave, which gives as reference, in the beginning, is converted a digital signal. In this study, PWM was used to convert the alternative voltage signal from the input to a digital signal that used to trigger the inverter switches. The basic single phase equivalent of PWM was given in Figure 5.

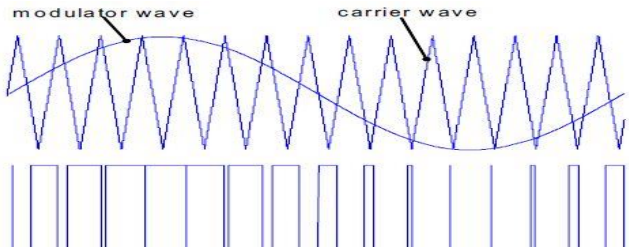

Figure 4
PWM technique [28]

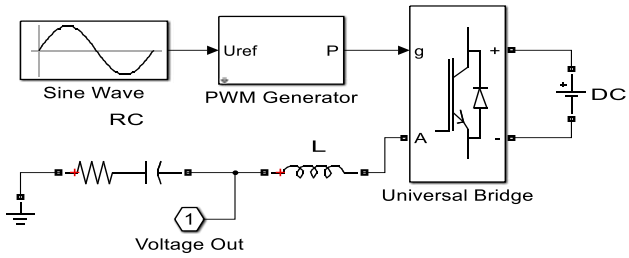

Figure 5
PWM generator block diagram

2.4 System Parameters

System parameters are determined as follows, the reactive power of the grid, the grid voltage, the amount of radiation in the environment and the active-reactive

power of the solar inverter. The general power formula is given in (1), the active and reactive power of the inverter are given in (2) and (3) respectively.

$$S = \sqrt{P^2 + Q^2} \tag{1}$$

$$P = V_i \frac{V_s}{2\pi f L_c} \sin\theta = P_{\max} \sin\theta \quad [29] \tag{2}$$

$$Q = \frac{V_i^2}{2\pi f L_c} - V_i \frac{V_s}{2\pi f L_c} \cos\theta \qquad [29] \tag{3}$$

In the above equations; S represents the apparent power, P inverter active power, Q inverter reactive power, Θ angle represents the phase angle between the grid voltage V_s and the inverter voltage V_i, the L coupling coil and the f frequency (50 Hz). In the active-reactive power control of the system, the phase angle with the inverter voltage amplitude holds an important place. Figure 6 shows that reactive power change versus the voltage change.

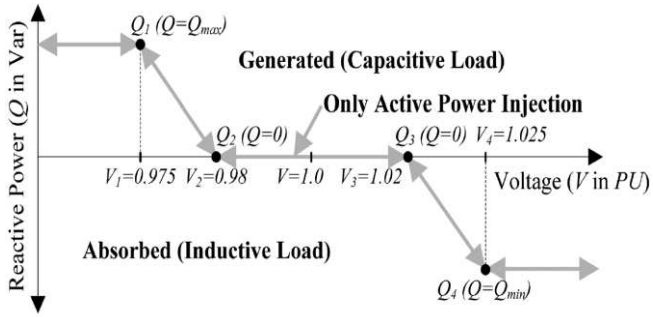

Figure 6
Reactive power-voltage change chart [30]

3 System Control

The parameters to be controlled in the system are the phase angle of the voltage at the output of the inverter to be connected to the grid, the amount of radiation in the environment, the reactive power of the inverter, the inverter output voltage amplitude and the grid voltage. Hysteresis band control was used to control these parameters. Hysteresis control works in case the specified parameters go beyond the limit values and return this parameter to the desired value by activating system control components. In this study, system control is done as follows; the voltage and reactive power values are continuously monitored by the sensors, the amount of radiation in the environment is also recorded. The measured voltage value is

compared to the nominal (reference) voltage value. If the difference between these two voltages is at the limit values and the amount of radiation in the environment is maximum, the inverter continues to generate only active power. If the difference between the reference voltage and the voltage measured exceeds the limit value, the inverter generates the reactive power with all its capacity to prevent the voltage drop. If there is a difference between the reference voltage and the grid voltage but not around the limit values and if the amount of radiation is not the maximum value, the inverter generates reactive power with the remaining capacity from the active power generation, depending on the state of the reactive power in the electric grid. This is done by controlling the phase angle and amplitude of the output voltage of the inverter. If the amount of radiation is zero (night mode), the inverter can use its full capacity for reactive power generation. In the night mode, the inverter is allowed to generate reactive power according to the grid voltage condition as in daylight mode. The logical flow diagram of the system is depicted in Figure 7.

4 Results

In this part of the study; various scenarios were tested in the Matlab/Simulink environment in order to see if the mentioned voltage regulation operation can be done with PV in the electrical grid where a solar system is connected and thus system performance was evaluated. In these scenarios, the amount of radiation in the environment, grid voltage and change of grid reactive power were modelled and how the inverter reacted to the grid in terms of active and reactive power was evaluated.

4.1 Case-1

In this scenario; it was assumed that the amount of radiation in the environment was 400 W/m^2 and the grid voltage decrease with the grid's inductive characteristic. However, this voltage drop was considered to be within the limit values. When Figure 8 shows the reactive power of the grid, Figure 9 shows the voltage amplitude in the case of this reactive power.

In the case of an inductive load is 916 VAR in the system, it is understood that the single-phase voltage amplitude of the grid is 305.7 V in Figure 8 and Figure 9. In this low voltage grid with a nominal value of approximately 310 V, there is a voltage drops approximately 4.3 V caused by both active and reactive loadings. The following graphs indicate that how much of this voltage drop was caused by inductive loading.

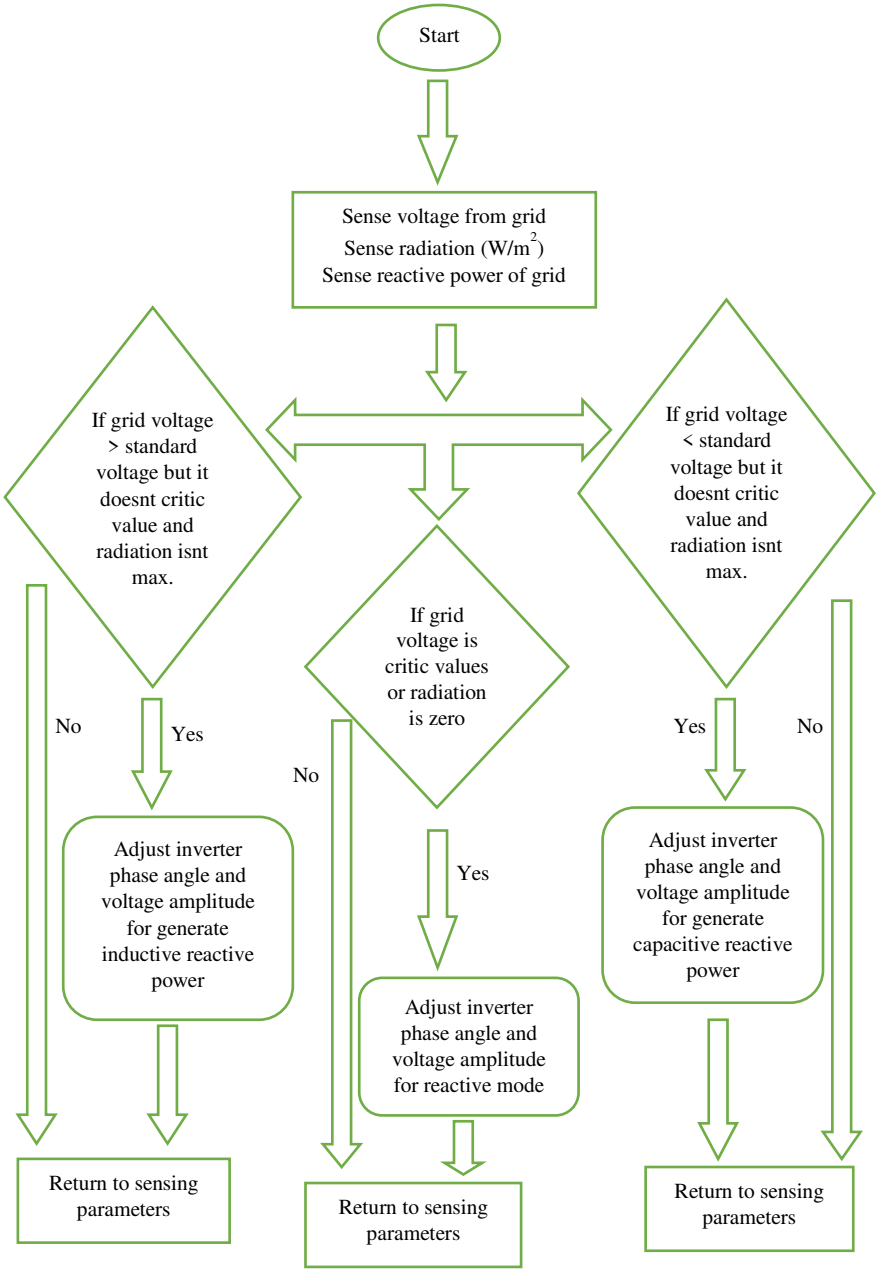

Figure 7
Block flow diagram of the system

The active-reactive power response of the 1 kVA solar inverter is shown in Figure 10. Figure 10 shows that the solar inverter was connected to the grid with a 1 second delay. This is due to the integration conditions required for frequency and voltage. In the 400 W/m^2 environment, the inverter uses it's 400 W capacity for active power generation, it is understood that the capacitive reactive power capacity of approximately -900 VAR of inverter was used to compensate the +900 VAR inductive reactive power where already exists in the grid.

Inverter and grid reactive powers were given together in Figure 11. From this graph, it can be seen that the inverter reacts by generating capacitive reactive power to the grid to eliminate the inductive load that exists in the grid.

It is observed in Figure 12 that the inductive reactive power is eliminated by the solar inverter and the grid voltage amplitude increases from 305.7 to 305.9 Volts. This 0.2 V voltage regulation is achieved with the selection of the inverter capacity as 1 kVA and the reactive power generation in the grid according to this selection. In other words, this effect will also increase in the selection of a larger grid and inverter.

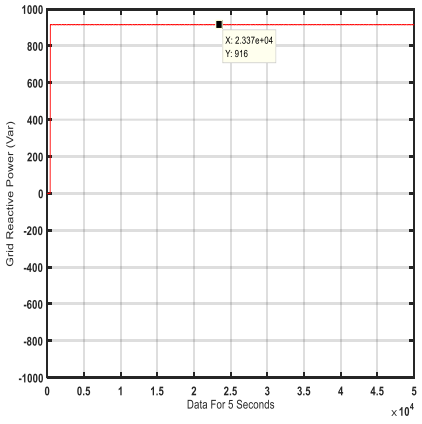

Figure 8
Inductive reactive power of grid

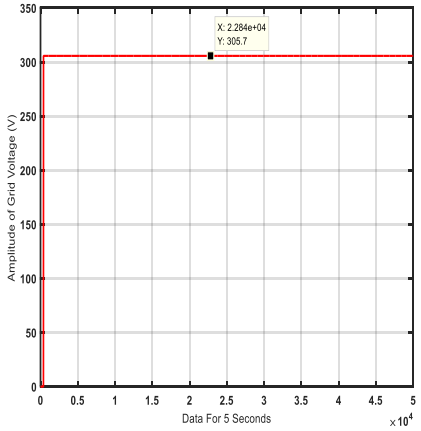

Figure 9
Grid voltage amplitude

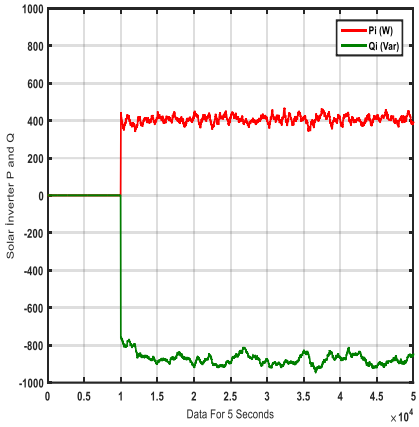

Figure 10
Active-reactive power response of inverter

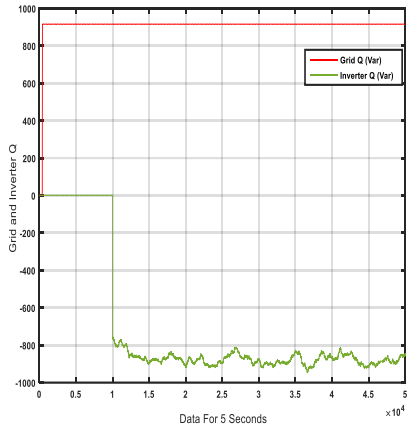

Figure 11
Grid and inverter reactive powers

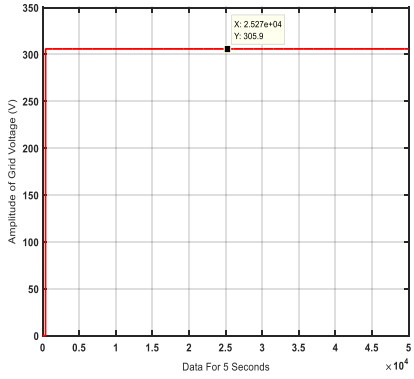

Figure 12
Regulated grid voltage

4.2 Case-2

In this scenario; approximately 800 W/m^2 of radiation amount exists in the midday and the situation where the voltage drop was close to the lower limit values (critical condition) was modelled. The system was designed to be inductive, and since the voltage drop was close to the limit value, regardless of the amount of radiation, the solar inverter used the entire power capacity for capacitive reactive power to pull the voltage higher and prevent the low voltage relays in the distribution grids from disabling the system. This means that the inverter does not generate active power notwithstanding the radiation amount. The critical value of the grid voltage amplitude and 1000 VAR inductive reactive power were given in Figure 13 and Figure 14 respectively.

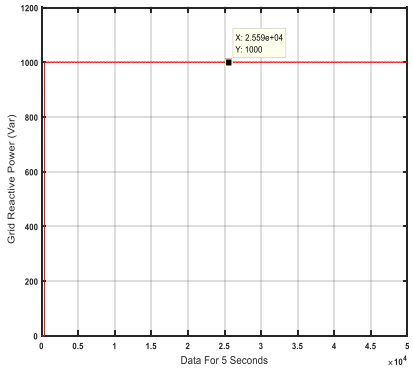

Figure 13
Inductive reactive power of grid

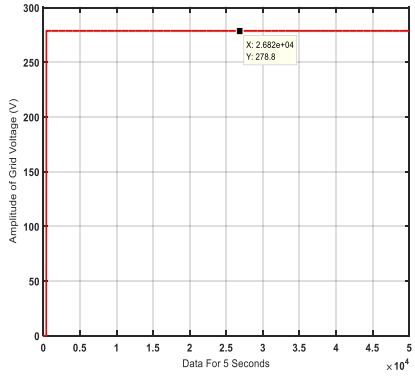

Figure 14
Grid voltage amplitude

Figure 14 shows that the grid voltage is 278.8 *V*, which is the critical voltage value due to the active and reactive power supplied by the system. In this case, the response of the inverter in terms of active and reactive power generation was given in Figure 15.

As seen in Figure 15, despite the fact that radiation is maximum, the inverter has used its entire capacity to generate capacitive reactive power as the grid voltage is at critical values.

As a result of this response of the inverter, the amplitude of the grid voltage was given in Figure 16. It was understood from the voltage amplitude graph of Figure 16 that initially, when the grid voltage was about to almost disable the grid with the amplitude of 278.8 Volts, the voltage amplitude was brought to 279 Volts through the capacitive response of the inverter thus it prevented the disable of grid.

In Figure 17, the reactive power of the grid and the inverter were given at the same time and so the reactive power contribution of the inverter to the grid was seen more clearly.

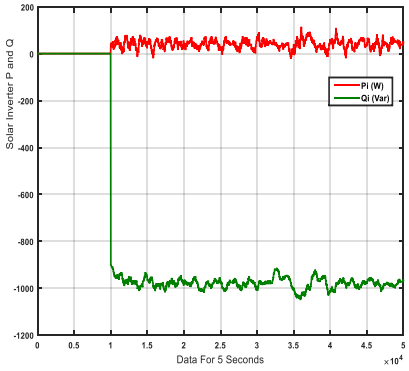

Figure 15
Active-reactive power response of inverter

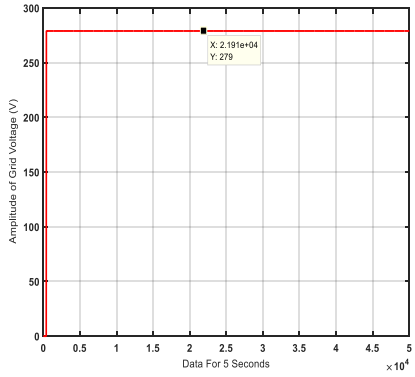

Figure 16
Regulated grid voltage

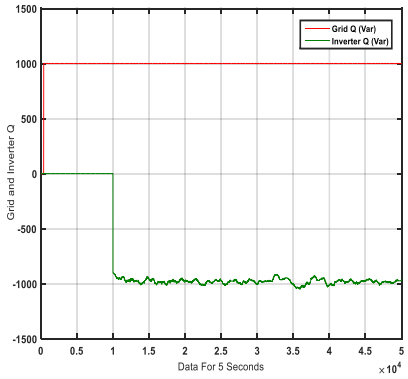

Figure 17
Grid and inverter reactive powers

4.3 Case-3

In this scheme of the study, it was taken into consideration that the amount of radiation is 600 W/m^2 and the grid has capacitive properties so that the grid voltage was increased but not critically. In this case, the response of the solar inverter and the changes in the grid voltage were evaluated. In this context, grid reactive power and grid voltage were given in Figure18 and Figure19 respectively.

Figure 18 shows that the capacitive reactive power of the grid is approximately -800 VAR. In Figure 19 the single phase voltage amplitude value is seen as 306 Volts in this grid which has capacitive character. The reaction of the solar inverter in terms of power balance was given in Figure 20. When Figure 20 is examined, it has been understood that the inverter consumes 600 *W* of capacity for active

power generation, the remaining power capacity of it is spent on inductive reactive power generation to suppress the capacitive reactive power in the grid.

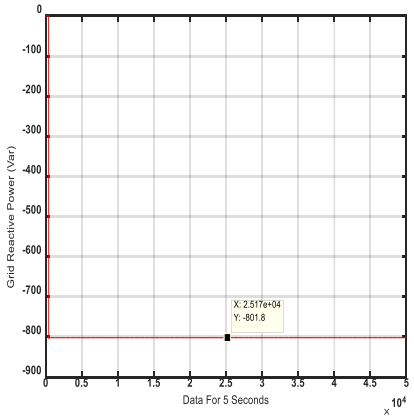

Figure 18
Capacitive reactive power of grid

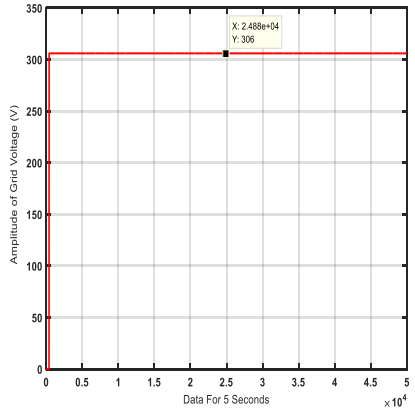

Figure 19
Grid voltage amplitude

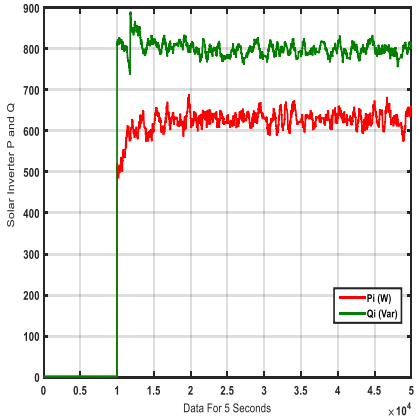

Figure 20
Active-reactive power response of inverter

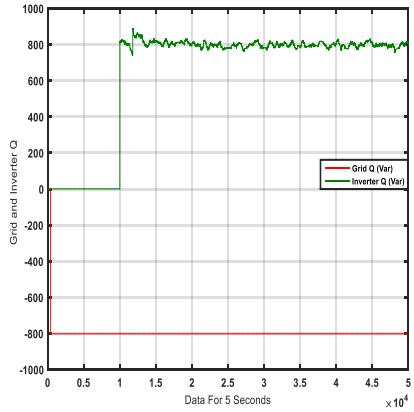

Figure 21
Grid and inverter reactive powers

The reactive power of the grid and solar inverter can be seen in the same coordinate of Figure 21. With this graph, the inductive reactive power response from the inverter to the existing capacitive reactive power in the grid can be understood more clearly.

In Figure 22, as a result of the reaction of the inverter to the grid, it is understood that the grid voltage amplitude value which is initially capacitive is decreased from 306 *V* to 305.9 *V* with only 800 VAR inverter intervention.

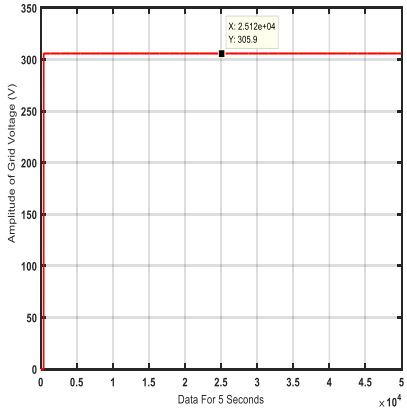

Figure 22
Regulated grid voltage

Conclusions

In this study, the aim was to improve system voltage by eliminating the reactive powers occurring in networks, by means of the opposite reactive powers and solar inverters connected to the electricity grid. The active and reactive power responses of these inverters to the grid were discussed by considering the presence of critical voltage in the network and the amount of radiation. In this context, the grid was modelled in Matlab/Simulink environment to see the system performance. Three different scenarios were performed in order to measure the response of the modelled system for different conditions. In the first scenario, an inductive grid was created and the condition where the grid voltage was not critical, was modelled. Accordingly, it was seen that the inverter contributes to the voltage adjustment by giving a capacitive response to the grid. In the second scenario; the grid was again considered inductively and the radiation amount was modelled to be high. However, with the assumption that voltage had a critical value, the inverter acted fully capacitive and it didn't generate active power to prevent grid voltage drop. In the last case; the grid was thought as capacitive and it was assumed that there is a certain amount of radiation in the environment. In this condition, it was seen that as the system voltage was not critical, the inverter contributes to the grid by generating active power up to the amount of radiation and generating inductive reactive power with its remaining power capacity.

As a result, it was understood, that reactive power generated in electrical grids can be eliminated by means of solar inverters and stable voltage regulation could be achieved in this way. It can be seen that the power losses can be reduced and the grids which were disabled, due to voltage drops, can continue to function. Eventually, good power quality, with an optimal and steady voltage and frequency supply, can be offered to consumers, through the elimination of the reactive powers in the grid.

Acknowledgment

We would like to thank to the Scientific Research Projects Unit of Kahramanmaras Sutcu Imam University, which contributes to doing this study with individual project which named as "Power Quality Improvement by Integration of Solar Systems into the Power Grid" and has **2017/7-202M** project number.

References

[1] Molina-García, Á., Mastromauro, R. A., García-Sánchez, T., et al.: Reactive power flow control for PV inverters voltage support in LV distribution networks. IEEE Transactions On Smart Grid, 2017, 8, (1), pp. 447-546 doi: 10.1109/TSG.2016.2625314

[2] Sadnan, R., Khan, Md. Z. R.: Fast real and reactive power flow control of grid-tie photovoltaic inverter. 9th International Conference on Electrical and Computer Engineering (ICECE), Dhaka, Bangladesh, December 2016, pp. 570–573, doi: 10.1109/ICECE.2016.7853984

[3] Safayet, A., Fajri. P., Husain, I.: Reactive power management for overvoltage prevention at high PV penetration in a low-voltage distribution system. IEEE Transactions On Industry Applications, 2017, 53, (6), pp. 5786-5794, doi: 10.1109/TIA.2017.2741925

[4] Islam, M., Afrin, N., Mekhilef, S.: Efficient single phase transformerless inverter for grid-tied PVG system with reactive power control. IEEE Transactions On Sustainable Energy, 2016, 7, (3), pp. 1205-1215, doi: 10.1109/TSTE.2016.2537365

[5] Wang, L., Yan, R., Saha, T. K.: Voltage management for large scale PV integration into weak distribution systems. IEEE Transactions on Smart Grid, 2018, (9), (5), pp. 4128-4139, doi: 10.1109/TSG.2017.2651030

[6] Kraiczy, M., Stetz, T., Braun, M.: Parallel operation of transformers with on load tap changer and photovoltaic systems with reactive power control. IEEE Transactions on Smart Grid, 2018, 9, (6), pp. 6419-6428, doi: 10.1109/TSG.2017.2712633

[7] Marrekchi, A., Kammmoun, S., Sallem, S., et al.: Sliding mode control of reactive power for three-phase grid-connected photovoltaic system. 6th International Conference on Systems and Control (ICSC), Batna, Algeria, May 2017, pp. 121-126. doi: 10.1109/ICoSC.2017.7958666

[8] Jabr, R. A.: Linear decision rules for control of reactive power by distributed photovoltaic generators. IEEE Transactions on Power Systems, 2018, 33, (2), pp. 2165-2174, doi: 10.1109/TPWRS.2017.2734694

[9] Jafarian, H., Parkhideh, B., Enslin, J., et al.: On reactive power injection control of distributed grid-tied AC-stacked PV inverter architecture. IEEE

Energy Conversion Congress and Exposition (ECCE), Milwaukee, WI, USA, September 2016, pp. 1-6, doi: 10.1109/ECCE.2016.7855303

[10] Sreekanth, T., Narasamma, N. L., Mishra, M. K.: A high gain grid connected single stage inverter system with reactive power control. IEEE International Conference on Industrial Technology (ICIT), Toronto, ON, Canada, March 2017, pp. 358-363, doi: 10.1109/ICIT.2017.7913257

[11] Liu, L. Y., Gao, J. T., Lo, K. Y.: A reactive power control strategy of the grid-connected inverter for microgrid application. IEEE 3rd International Future Energy Electronics Conference and ECCE Asia (IFEEC - ECCE Asia), Kaohsiung, Taiwan, June 2017, pp. 755-759, doi: 10.1109/IFEEC.2017.7992134

[12] Li, D., Ho, C. N. M., Liu, L., et al.: Reactive power control for single-phase grid-tie inverters using quasi sinusoidal waveform. IEEE Transactions on Sustainable Energy, 2018, 9, (1), pp. 3-11, doi: 10.1109/TSTE.2017.2710340

[13] Anurag, A., Yang, Y., Blaabjerg, F.: Thermal performance and reliability analysis of single-phase PV inverters with reactive power injection outside feed-in operating hours. IEEE Journal of Emerging and Selected Topics in Power Electronics, 2015, 3, (4), pp. 870-880, doi:10.1109/JESTPE.2015.2428432

[14] Liu, J., Li, Y., Liu, F., et al: An improved reactive power control strategy for overvoltage fluctuation of distribution network with high penetration of PVs. 36th Chinese Control Conference (CCC), Dalian, China, July 2017, pp. 9321-9325, doi: 10.23919/ChiCC.2017.8028842

[15] Jain, C., Singh, B.: Single-phase single-stage multifunctional grid interfaced solar photo-voltaic system under abnormal grid conditions. IET Generation, Transmission & Distribution, 2015, 9, (10), pp. 886-894, doi: 10.1049/iet-gtd.2014.0533

[16] Akutsu, H., Hirata, K., Ohori, A., et al.: Decentralized active and reactive power control for PV generation plants using real-time pricing strategy. American Control Conference (ACC), Seattle, WA, USA, May 2017, pp. 2761-2766, doi: 10.23919/ACC.2017.7963369

[17] Zhang, F., Guo, X., Chang, X., et al.: The reactive power voltage control strategy of PV systems in low-voltage string lines. IEEE Manchester PowerTech, Manchester, UK, June 2017, pp. 1-6, doi: 10.1109/PTC.2017.7980995

[18] Pothisoonthorn, S., Ngamroo I., Kunakorn, A.: Power swing and voltage stablization by PV generator with active and reactive power controls. IEEE Manchester PowerTech, Manchester, UK, June 2017, pp. 1-6, doi: 10.1109/PTC.2017.7981073

[19] Liu, L., Gao J., Lo, K. Y.: A reactive power control strategy of the grid-connected inverter for microgrid application. IEEE 3rd International Future Energy Electronics Conference and ECCE Asia (IFEEC - ECCE Asia), Kaohsiung, Taiwan, June 2017, pp. 755-759, doi: 10.1109/IFEEC.2017.7992134

[20] Zeng, R., Wang Z., Chinthavali, M. S.: Coordinated reactive power control of single-phase grid-connected converter for distributed energy resources integration considering multiple functionality conflicts. IEEE Transportation Electrification Conference and Expo (ITEC), Chicago, IL, USA, June 2017, pp. 747-753, doi: 10.1109/ITEC.2017.7993363

[21] Behera, M. P., Ray1, P. K., Beng, G. H.: Single-phase grid-tied photovoltaic inverter to control active and reactive power with battery energy storage device. IEEE Region 10 Conference (TENCON), Singapore, Singapore, November 2016, pp. 1900-1904, doi: 10.1109/TENCON.2016.7848352

[22] Acikgoz, H., Kececioglu, O. F., Gani, A., et al.: Robust control of shunt active power filter using interval type-2 fuzzy logic controller for power quality improvement. Technical gazette, 2017, 24, Supplement (2), https://doi.org/10.17559/TV-20161213004749

[23] Biel, D., Scherpen, J. M. A.: Passivity-based control of active and reactive power in single phase PV inverters. IEEE 26th International Symposium on Industrial Electronics (ISIE), Edinburgh, UK, June 2017, pp. 999-1004, doi: 10.1109/ISIE.2017.8001382

[24] Pal, D., Bajpai, P.: Active and reactive power control in three phase solar PV inverter using modified IC method. 21st Century Energy Needs - Materials, Systems and Applications (ICTFCEN), Kharagpur, India, November 2016, pp. 1-6, doi: 10.1109/ICTFCEN.2016.8052727

[25] Sekkeli, M., Tarkan, N.: Development of a novel method for optimal use of a newly designed reactive power control relay. Electrical Power and Energy Systems, 2013, 44, (1), pp 736-742, https://doi.org/10.1016/j.ijepes.2012.08.015

[26] Hassan, F., Critchley, R.: A robust PLL for grid interactive voltage source converters. Proceedings of 14th International Power Electronics and Motion Control Conference EPE-PEMC, Ohrid, Macedonia, September 2010, T2-29-T2-39, doi: 10.1109/EPEPEMC.2010.5606822

[27] Jalilian, M., Nouri, M., Ahmadi, A., et al.: Pulse width modulation (PWM) signals using spiking neuronal networks. IEEE International Conference on Signal and Image Processing Applications (ICSIPA), Kuching, Malaysia, September 2017, pp. 180-184, doi: 10.1109/ICSIPA.2017.8120602

[28] Gutiérrez, M. J. M., Gonzalez, A. R., Merino, F. V.: Low heating losses by harmonic injection PWM with a frequency modulated triangular carrier.

International Symposium on Power Electronics, Electrical Drives, Automation and Motion, Ischia, Italy, June 2008, pp. 1398-1401, doi: 10.1109/SPEEDHAM.2008.4581229

[29] Albuquerque, F. L., Moraes, A. J., Guimara˜es, G. C.: Photovoltaic solar system connected to the electric power grid operating as active power generator and reactive power compensator. Solar Energy, 2010, 84, (7), pp 1310-1317, doi: org/10.1016/j.solener.2010.04.011

[30] Kim, I.: Optimal distributed generation allocation for reactive power control. IET Generation, Transmission & Distribution, 2017, 11, (6), pp. 1549-1556, doi: 10.1049/iet-gtd.2016.1393

Design and Construction of a Sun Simulator for Laboratory Testing of Solar Cells

István Bodnár[1], Dániel Koós[2], Patrik Iski[3], Ádám Skribanek[4]

[1,2,3,4]University of Miskolc, Faculty of Mechanical Engineering and Informatics,
[1,3,4]Isntitute of Electrical and Electronic Engineering,
[2]Isntitute of Energy Engineering and Chemical Machinery,
[1]vegybod@uni-miskolc.hu, [2]elke100@uni-miskolc.hu, [3]elkvmt@uni-miskolc.hu,
[4]elkskrib@uni-miskolc.hu, [1,2,3,4] H-3515 Miskolc-Egyetemváros

Abstract: The current study is dedicated to the realization and development of an ASTM E972 standardized, class C sunlight simulator. Sunlight simulators or sun simulators are laboratory devices used to study solar cells. Our sun simulator is suitable for the examination and qualification of solar cells, with an area of 150 mm x 150 mm. It consists of 4 LED blocks of different colors and a halogen reflector. The completed sun simulator is validated; the design and realized parameters display good matching.

Keywords: sun simulator; solar simulator; solar cell tester; halogen; LED

1 Introduction

Energy production, which inevitably involves pollution, is indispensable in our society. Producing energy in a way that is environmentally-friendly is possible with renewable sources, of which, the source is usually (with some exceptions, such as geothermal energy and nuclear) solar energy related. This energy reaches the surface in the form of radiation, which can be harvested via solar panels, solar thermal collectors and/or solar concentrators.

Today's widespread mono- and polycrystalline solar panels work with an efficiency of 14-18%. Research is continuously being carried out, in order to achieve a higher rate of efficiency and complex investigations of currently used technologies are also required. The quality of illumination of solar panels is a high priority factor in laboratory examinations, which is determined by how similar the facilitated sun simulator's light is compared to natural illumination [5, 8, 21, 24].

A sun simulator is a device that provides illumination, of which, the intensity and spectral composition is similar to that of natural sunlight inside the inspection range. The similarity depends on the light of the facilitated sun simulator. Using

such equipment, the simulated sunlight is continuously available, relatively homogeneously distributed, hence can be used for various indoor or long-term investigations. Since it is regulable, it is commonly used to inspect solar cells, awnings, plastics and other materials, and sunlight-sensitive devices [5, 21, 24].

Sun simulators usually consist of three main units. Any sun simulator has light sources and power supplies. Optical light filters can be used to achieve appropriate illumination, and the use of a control unit is essential so that the properties of the light can be changed as desired.

First generation sun simulators were generally based on halogen, xenon, or tungsten light sources, and their combinations. To achieve homogeneous light intensity, complex optical systems had to be installed. Acquiring the desired spectral composition required optical filters. In the 1990s, with the rapid advancement of semiconductor technology, new sun simulator types have seen the light. They operated with a cheaper, more compact and more energy efficient alternative of previously mentioned light sources; LED units. These devices can be described as easily regulable, and a huge advantage of LED based light sources is that, by applying different colors, a similar spectral composition to that of natural sunlight can be achieved. Note that by relying solely on LED lights, 1,000 W/m^2 light intensity, which is required for standardized examinations, can hardly be achieved. However, applying a combination of LED lights and conventional illumination can mean the solution to this problem [1-4, 12, 25].

Our goal is to develop an ASTM E972 standardized [1-4, 12, 25], class C sun simulator, which is suitable to illuminate an inspection area of 150 mm x 150 mm, making the standardized examination of small size solar cells possible.

2 Design

This section briefly describes the relevant standards, and details the steps and results of the design process.

2.1 ASTM E972 Standard

Requirements considering sun simulators are found in the *American Standard for Testing and Materials* (ASTM) E972. According to this standard, sun simulators can be classified by three main aspects into different classes, however, any sun simulator must be able to provide the average surface light intensity (AM 1.5) with a maximal 1,000 W/m^2 value [1-4, 25].

When inspecting the conformity of a sun simulator with standards, the similarity between the spectral composition of the illumination and natural sunlight is examined, as well as the temporal and spatial unevenness of said illumination.

The standard states that the spectral conformity in case of AM 1.5 is to be examined between the wave lengths of 400 and 1,100 nm. Sun simulators can then be classified into class A, B or C based on the previously mentioned three aspects. The requirements of the three groups are summarized in Table 1 [1-4, 25].

Table 1
Category requirements for standardized sun simulators

	Class A	**Class B**	**Class C**
Spectral conformity with natural sunlight	75-125%	60-140%	40-200%
Spatial unevenness	≤ 2%	≤ 5%	≤ 10%
Temporal unevenness	≤ 2%	≤ 5%	≤ 10%

The stricter requirements of class A and B sun simulators carry a higher cost and an increased time of development, hence taking the feasibility into account, we set the development of an ASTM E972 standardized, class C sun simulator as the main objective for our research and development project. For future research projects with small size solar cells, said class C simulator provides a suitable laboratory light source, as the solar intensity to be provided is the same for every class. For a class C sun simulator, however, the deviation of spectral conformity, and its spatial and temporal unevenness hardly affect the final measurement results, therefore the use of a higher-class sun simulator is unnecessary to successfully carry out the measurements [1-4, 25].

2.2 Design Parameters, Construction Basics

Our goal is to design a class C sun simulator which complies the ASTM E972 standard. The sun simulator consists of a combination of high-performance, colored LED units and halogen lamps; by implementing LED units with appropriate colors, a similar spectral composition to that of natural sunlight can be achieved, that is between the wave lengths of 400 and 700 nm. But relying solely on LED lights, the 1,000 W/m^2 light intensity required for standardized examinations can – in this case – not be achieved, hence, the application of halogen lamps. Further advantages of the concept are that by arranging the LED units in a matrix form, a relatively homogeneous light intensity distribution can be achieved, and the spectral distribution of the halogen lamp is similar to that of the sunlight within the wavelength range of 700 and 1,100 nm [1-4, 9, 12, 14, 25].

After laying the foundations of the construction, it is now necessary to define a number of parameters which are going to significantly influence future results.

In our case, we can distinguish fixed parameters and so-called design parameters. Fixed parameters are set based on previous measurements or considerations, while the optimal values of the design parameters will be determined in a later phase of the design process. Fixed parameters are summarized in Table 2.

Table 2
Fixed parameters of LED units, halogen light sources, and the inspection range

LED	
Number	36 pieces (6 x 6 matrix arrangement)
Colours	neutral white, red, green, blue
Distance between LED units in the matrix	3 cm
Maximal performance of LED units	3 W
Halogen lamps	
Maximal performance of halogen lamps	50 W
Placements of halogen lamps	at the corners and/or sides of the LED matrix
Number of halogen lamps	2 x 4 = 8 pieces
Inspection range	
Size of inspection range	15 cm x 15 cm
Distance from the LED matrix	8 cm

The size of the inspection range is determined by the size of the solar cells, which is usually 15 cm x 15 cm. The number of LED units can be determined by considering that the size of the matrix is slightly larger than the size of the inspection range, but it is also necessary for the final number of LEDs to be a multiple of four, since four LED colors are implemented. The main considerations when choosing the colors of the LED units are the availability and the supposed spectral composition of individual colors. The available color palette is as follows: neutral white, cold white, orange, red, green, blue. The cold white spectrum is similar to the neutral white, but it has a dominant blue vibe, which becomes dispensable by implementing the blue as an individual color. The spectrum of orange highly overlaps with red, hence using both at the same time is superfluous. Accordingly, neutral white, red, green, and blue are chosen.

The distance between the LED matrix and the inspection range is chosen to be as short as possible to maximize light intensity, but taking the halogen lamps into account as well, which due to their size require a larger space, 8 cm distance between the two units is necessary. Early in the design, the question of the adjustability of light intensity of the halogen lamps arose. Therefore, a high performance (50 W, GU10 socket) halogen lamp seems to be the most expedient choice, giving us the opportunity to adjust the light intensity on a wide scale.

Based on preliminary light intensity measurements, we found that 4 pieces of 50 W halogen lamps may not be enough to provide the necessary 1,000 W/m^2 light intensity. The number of halogen bulbs is 4 or 8 pieces (which will further be unfolded later in the optimization part). Keeping the symmetry in mind, halogen lamps can be placed in the corners and/or on the sides.

Contrary to the fixed parameters, some variable parameters require a more detailed design. For LED units, such parameters are the supply amperage of individual color groups. In case of the halogen lamps, we are looking for the optimal number (4 or 8 units) and their positions (horizontal inclination, horizontal and vertical distance from the inspection range).

2.3 Designing for Spectral Conformity

One criterion required by the standard, is the spectral conformity of the illumination with sunlight. When designing for spectral conformity, the first step is to determine the spectrum curves of the light sources. However, during the design process we do not have access to a spectrometer to perform an absolute spectral assay, so the spectrum curves are obtained indirectly. Calculations are based on light intensity test measurements, calculations and evaluations are done in MATLAB environment. (The MATLAB code can be found at: http://www.uni-miskolc.hu/~vegybod/sunsimulator.m).

In case of the LED units, the spectrum curves can be approximated by Gaussian curves. For this, three practically chosen wavelength - light intensity points are necessary, to which the desired curve will be approximated by the MATLAB code. Similar algorithms are used by Hussain et al. [25]. For LED units, the wavelength values of emission peaks for each colour are given by the manufacturer, and for each colour, the wavelength range in which the emission is still significant is also known. Therefore, we have the necessary points. Hereupon, the relative spectrum curve becomes obtainable for the given colour, although for design purposes the absolute spectral distribution is necessary. This calibration problem can be resolved by considering that the integral of the absolute spectrum curve of the given light source is approximately the same as the value of light intensity it produces. So, if the light intensity and amperage of the different LED colours are known for a particular arrangement, then all the information that is necessary to determine the absolute spectrum curve is given.

In case of halogen lamps, the absolute spectrum structure can be derived from the black-body radiation, assuming a color temperature of 3,200 K using equation 1 [5-8, 14, 16, 23].

$$E(\lambda,T)=\frac{2\pi hc^2}{\lambda^5\left(e^{\frac{hc}{\lambda kT}}-1\right)} \tag{1}$$

where: E – spectral intensity
λ – wavelength of light
T – color temperature
h – Boltzmann-constant
c – the velocity of light in vacuum

The superposition of the spectrum curves of individual light sources gives the specific spectrum structure of the sun-simulator for the given variables. The next task is to determine the amperage that can provide the best possible spectral conformity and complies the standard to power the LED units. The spectral conformity is calculated by equation 2 – in case of AM 1.5 global emission – in the wavelength range 400 - 1,100 nm [5-8, 14, 16, 23].

$$SE(\lambda_a - \lambda_f) = \frac{\int_{\lambda_a}^{\lambda_f} E_{NSz}(\lambda)d\lambda}{\int_{\lambda_a}^{\lambda_f} E_{AM\,1,5}(\lambda)d\lambda} \tag{2}$$

where: $SE(\lambda_a - \lambda_f)$ – Spectral difference for the given wavelength range,
λ_a – Lower limit of the wavelength range
λ_f – Upper limit of the wavelength range
E_{NSz} – Spectral intensity of the sun simulator
$E_{AM\,1.5}$ – Spectral intensity of the sun

The maximum amperage capacity of the 3 W LED modules is 700 mA.

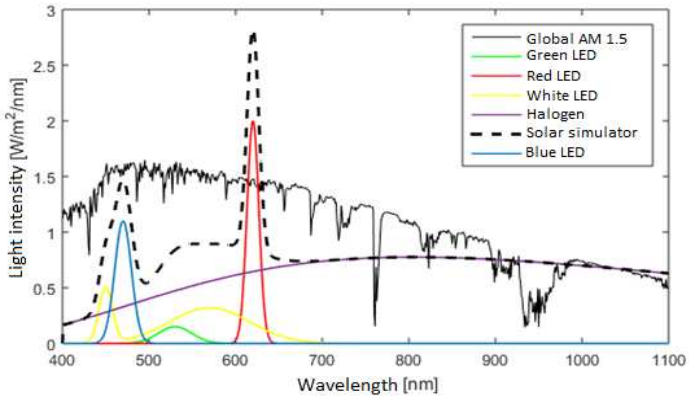

Figure 1
Spectral composition in case of the best matching LED-halogen combination

Five cases with different amperage values are investigated for every color by the MATLAB program we have written, starting from 700 mA decreasing by 100 mA steps. For the four colors, this means 5^4= 625 combinations in total. Thus, the calculation is executed for 625 amperage combinations; individual spectrum curves and, after the superposition, the specific spectrum structure of the sun-simulator is determined, and the spectral conformity is calculated in every case. The LED power supply combination with the best matching spectrum conformity

can be selected from the results of the iterations [1-4, 16, 25-29]. The spectral compositions of individual light sources, the sun simulator, and natural sunlight in case of the best matching amperage combination are shown in Figure 1 [5-8, 14, 16].

2.4 Designing for Light Intensity Distribution

Prior to the proper design of the light intensity distribution, test measurements take place. The light intensity distribution generated by the LED and the halogen illumination is calculated separately, and their superposition gives the light intensity distribution produced by the sun simulator. For LED units, measurements are made with the power amperage values used for spectral matching.

The test range is divided into 3 cm x 3 cm squares and the light intensity measurement is performed in each square. Such a measurement results the light intensity distribution generated by a single LED unit for the given amperage. The expected light intensity distribution is calculated in MATLAB for the array of LED units' matrix, and the calculation is run for 625 amperage combinations [16]. Regarding the halogen lamps, variable parameters are their number (either 4 or 8) and their positions. Tests of 30 combinations of halogen position – number pairs are made. The halogen lamp can be positioned at the corners or at the sides of the test range, similarly to the previous one, and the light intensity measurement is performed again in each small square. The result is a light intensity distribution formed by a single halogen lamp at a given position, from which the future light intensity distribution will be determined. Figure 2 illustrates the light intensity distribution obtained by LED and halogen light sources [17-19].

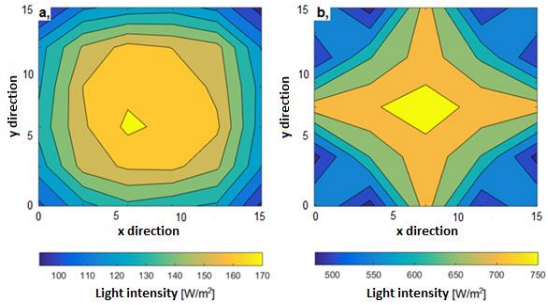

Figure 2
Light intensity distribution of (a) LED matrix, (b) halogen lamps

After the superposition of light intensity distributions produced by the LED and halogen lamps, the next step is to find the variables, with which the inhomogeneity of the light distribution is at nadir. According to the correlation

given by the standard, the inhomogeneity of the light distribution can be calculated by:

$$T_{Egy} = \frac{(I_{max} - I_{min})}{(I_{max} + I_{min})} \cdot 100\% \tag{3}$$

where: T_{Egy} – unevenness of light intensity
E_{max} – maximal light intensity
E_{min} – minimal light intensity

The upcoming process is similar to the method used to obtain the spectral distribution. The expected light intensity distribution is calculated for the 625 amperage and 30 position combinations by the MATLAB program, then the inhomogeneity of the current distribution is calculated.

The variable parameters that provide the highest homogeneity can be selected from the results of individual iterations. Figure 3 shows the light intensity distribution with the highest homogeneity. (Note: changing the light intensity of LEDs slightly affects the final light intensity distribution, the positioning of the halogen lamps has a much stronger impact.)

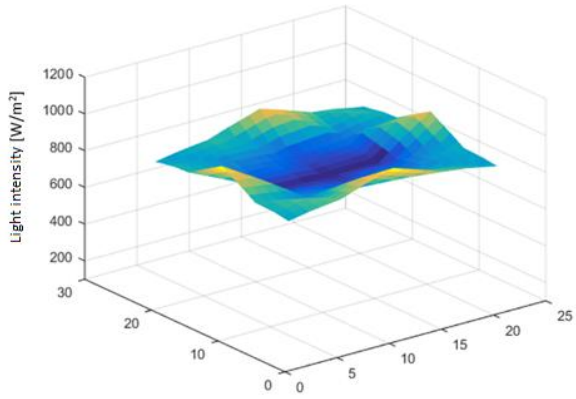

Figure 3
Light intensity distribution with the highest homogeneity inside the inspection range

2.5 Results of the Design Process

The purpose of the design method is to determine the optimal values of the previously described variable parameters (the power supply amperage of LEDs, quantity and position of halogen lamps), and to outline the actual design of the device to be constructed. Our designed sun simulator consists of a combination of 36 pieces of colored, 3 W LED lights, and 8 pieces of 50 W halogen lamps. Table 3 summarizes the power input of each LED colors separately. (When determining

these parameters, both spectral conformity and proper light intensity distribution are taken into account.) Table 4 summarizes the ideal positions of the halogen lamps.

Table 3
Power input values for different LED colors

Color	**I [mA]**	**U [V]**	**Color**	**I [mA]**	**U [V]**
Red	710	2.5	Green	670	3.2
Blue	710	3.4	White	680	4.0

Table 4
Planned positions of the halogen lamps

Horizontal distance from the LED matrix	Corner: 30mm	Side: 34mm
Horizontal inclination	Corner: 15°	Side: 45°
Vertical distance from the inspection range	Corner: 75mm	Side: 60mm

The effectiveness of the design phase can be evaluated according to compliance to standard. In the case of optimum variable parameters, the inhomogeneity of the projected sun simulator's light intensity is 12% based on the calculations and the average light intensity distribution at this time is 910 W/m^2. The degree of spectral matching in each wavelength range is given in Table 5.

Table 5
Calculated spectral conformity within the inspection range [5-8]

Wavelength range	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1.1
Spectral conformity	12.8	67.0	82.3	67.9	81.5	115.3

According to the standard conformity, the following conclusions can be drawn:

- Light intensity and the homogeneity of the light is slightly below of what is specified by the standard. This problem can likely be solved by further optimizing the positions of the halogen lamps on the physical framework.
- As for spectral conformity, the illumination meets the requirements of the class C standard at every wavelength range, except for the 400nm to 500nm interval. This result suggests the necessity to incorporate a new, low-wavelength dominant (ultraviolet) LED unit.

3 Construction and Realization

This chapter describes the construction of the sun simulator with the previously detailed parameters.

3.1 Power Supply

Power supply can be divided into two groups, since the power supply of the halogen lamps and the LED units must both be insured. Powering the halogen lamps is rather simple, since the 8 halogen lamps (with GU10 socket) are powered from the electrical grid through two dimmer switches [12, 17-19]. The halogen lamps can further be divided into two groups according to their dimming; these groups are halogen lamps on the sides and in the corners. Since the dimming of the groups is separated, the load on any dimmer will never exceed 4 x 50 W = 200 W. Our specific dimmer has an operation range between 40 W - 400 W, TRIAC based phase-splitting power control. Bliss et al. [3] and Kim et al. [4] show solutions for this purpose, where they combine LEDs with solar halogens.

For LED units, the goal is to control the brightness of each individual color groups. To ensure the feasibility of future fine tuning, higher power (5 W) LED units are installed. LED units of the same color are connected in series and are supplied by individual power supplies through individual power controllers. The actual LED matrix wiring plan is shown in Figure 4.

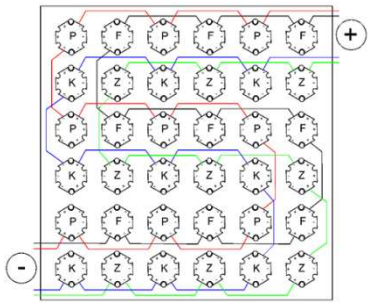

Figure 4
Positions and wiring of the LED units in the matrix

In case of LED lighting, it is essential to keep the amperage and voltage limit, which – in the case of these 3 W LED units – is 700 mA and 4 - 4.5 V. Given these parameters, the color groups of the LED matrix are powered through a combination of four dimmable current sources and power regulators of the same type as the halogen lamps connected to the AC side. This design is capable of supplying the system with the maximal voltage supply, and can dim the light intensity through a phase-splitting based controller.

To ensure the proper operation conditions of the high-performance LED units, the heat dissipation of the device must be absorbed somehow, which in practice, is most often solved by installing a heatsink. In our case the heat sink must fulfil two roles; first, it must have the appropriate heat transfer properties to be able to dissipate the generated heat. Second, it must have a large enough useful surface area that it can hold the designed LED matrix [1-4, 12]. At the beginning of the

design process of the heat dissipation system, we suspect that heat sinks with an affordable price will not be able to provide the necessary heat dissipation rate, hence we decided to incorporate a fan. Further calculations are done assuming the presence of forced convection. Geometric dimensioning is done in an inverse method, meaning that some heat sink and fan combinations are selected, and then checked individually. After that, we choose the one that is most suitable for our demands. From here on out, the verification of the chosen heatsink – fan combination will be unfolded. The material of the chosen heat sink is aluminum, its bounding size is 190 mm x 190 mm x 50 mm [length x width x height]. Active cooling is provided by a fan with a diameter of 140 mm. The nominal airflow rate of the fan is 126 m^3/h. The first step of the verification is to determine the thermal resistance of the heatsink given a certain airflow rate, which is calculated in an online program called *My Heat Sink- Thermal Resistance Calculator* [5]. After specifying the geometrical dimensions and the material of the heatsink, we get the thermal resistance as the function of the airflow rate.

According to relevant literature, the semiconductor material of the LED lights will remain properly operable below 125 °C [7, 15, 22]. In our case, the chosen heat sink – fan combination guarantees the proper operation of the LED units even at maximum loads the temperature is 88.9 °C [5, 12].

The last step is to provide the power supply of the cooling fan. The selected fan requires 12 VDC to operate, which is provided by a 12 V, 2 A, 24 W industrial power supply unit. The entire power supply wiring of the sun simulator is shown in Figure 5.

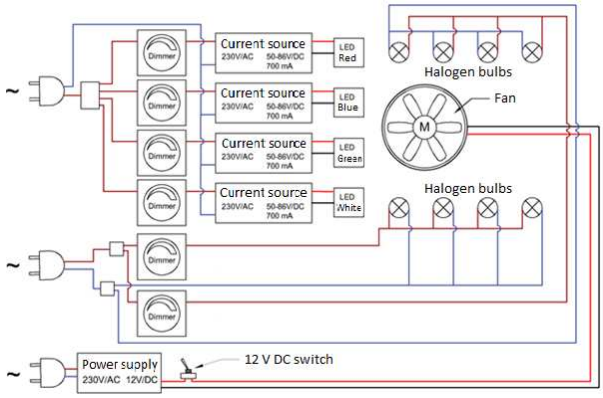

Figure 5
The power supply wiring of the sun simulator

3.2 Framework

The following main aspects are considered when designing the framework:

- proper holder for the heatsink with the LED matrix
- stable halogen framework that can provide positional adjustability
- no interference with the light sources (by casting shadows)

The material of choice is an iron hollow section with a cross section of 20 mm x 20 mm, which is easy to weld, has high bearing capacity and is since quite stable. It is not liable to warping or deformation, and besides can be used to construct a relatively slim structure. Halogen bulbs are held by threaded stems attached to the grooves of radially placed grooved flat iron rods [17-19]. The heatsink hangs from four threaded iron balls. Above located the support bracket of fan providing the active cooling is. The horizontal, "H" shaped part, to which all the other elements mentioned so far connect, can be considered the central unit of the structure. The vertical elevation of said unit is provided by four feet, while the horizontal crosses on the feet are required to achieve high stability, since the center of mass is located comparatively high. The inspection range is located below the heatsink of the LED matrix [12]. The height of the inspection range is planned to be adjustable. The bounding size of the framework is 600 mm x 550 mm. The visual representation of the completed framework is shown in Figure 6.

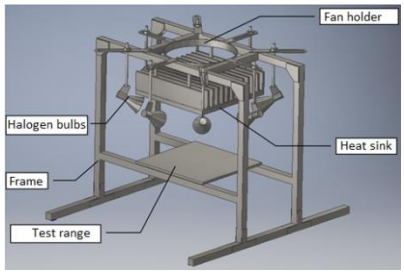

Figure 6
The designed construction of the sun simulator, with the main elements indicated

3.3 The Realized Construction

The Realized Construction can be seen in Figure 7. A two-channel laboratory power source with two current limiters are used to power the LED matrix, as shown in Figure 8.

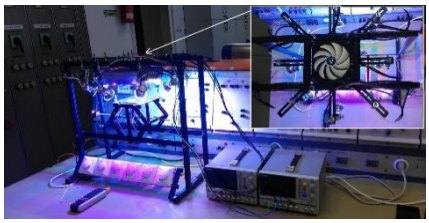

Figure 7
The complete sun simulator

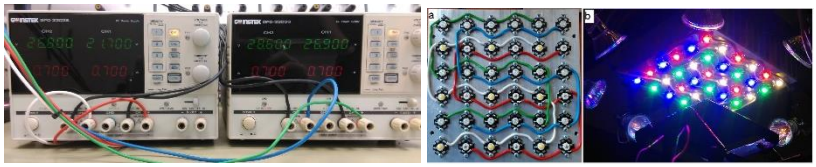

Figure 8
Power supplies used for the validation measurements and (a) Wiring of the LED matrix, (b) LED matrix during operation

4 Validation of the Design

This chapter is bound to describe the measurement of the previously calculated values (spectral structure, light intensity distribution, temperature of the LED), hereby describing the validation process of the sun simulator [12, 13].

4.1 Inspection of Light Intensity Distribution

An important quality of standard sunlight simulators is the magnitude and the distribution of the light intensity measured in the illuminated range. A light intensity measurement is performed on the finished construction. The inspection range – a square with width and depth of 15 cm – is split up to 1 cm^2 segments, then the light intensity is measured by a handheld light intensity [W/m^2] meter in every segment [20]. Measurements are done as described in Figure 9. As opposed to pre-measurements, this time the illumination is provided by the sun simulator with the previously defined fixed and design parameters. The resulting light intensity distribution is shown in Figure 10.

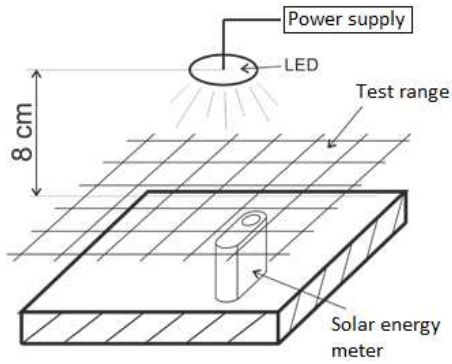

Figure 9
Measurement setup for measuring light intensity distribution

By examining the measured values, it can be deducted that light intensity values are within a certain range, depending on the measuring position. The smallest measured intensity value is Imin = 868 W/m^2 and the largest is Imax = 1,060 W/m^2. The arithmetic mean of the measured light intensity values is 951 W/m^2. It can clearly be seen that the light intensity distribution is relatively homogenic within the inspection range. Inhomogeneity can be calculated by a standardized equation (3) [20].

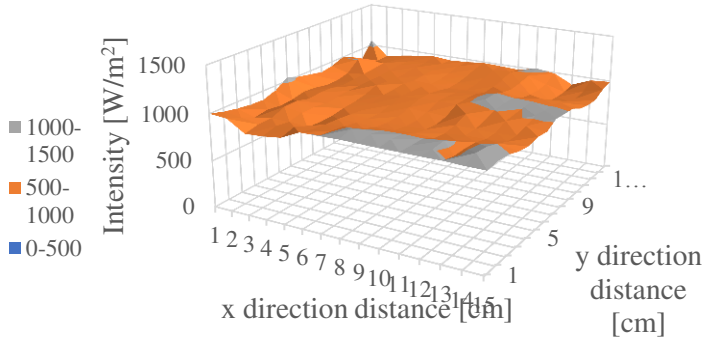

Figure 10
Measured light intensity distribution within the inspection range

According to the standard, for class C class sun simulators, the inhomogeneity of the light intensity distribution may not be more than 10%. The light intensity distribution of our sun simulator has a 9.96% inhomogeneity, which means the device complies with the standard.

4.2 Inspection of Spectral Structure

To check the spectral structure of the illumination, an *Ocean Optics USB 4000* type spectrometer is used. Due to the high directional dependence of the device, a white sheet of paper is placed atop the inspection range, and the probe is fixed in a position facing the scattered light. The measurement environment is shown in Figure 11.

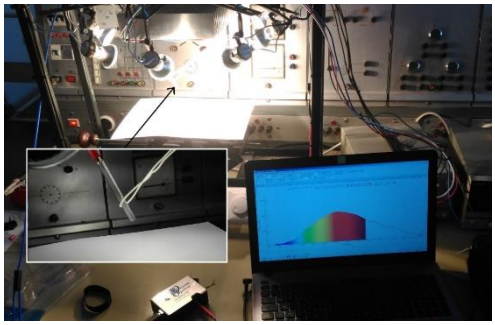

Figure 11
Spectral structure measurement

During the spectrum analysis, the spectrum structure of the finished sun simulator is examined. The applied spectrometer is not suitable to perform absolute light intensity measurements; however, it allows for the spectral curve of the sun simulator and natural sunlight to be recorded. Thus, it is possible to compare the spectrum structure of the sun simulator and natural sunlight, and calculate their spectral conformity [5-8, 20]. The measured spectral curves of the sun simulator and natural sunlight are shown in Figure 12. The same tendency can be seen in some similar publications, such as Kádár et al [8]. and Kenny et al. [20].

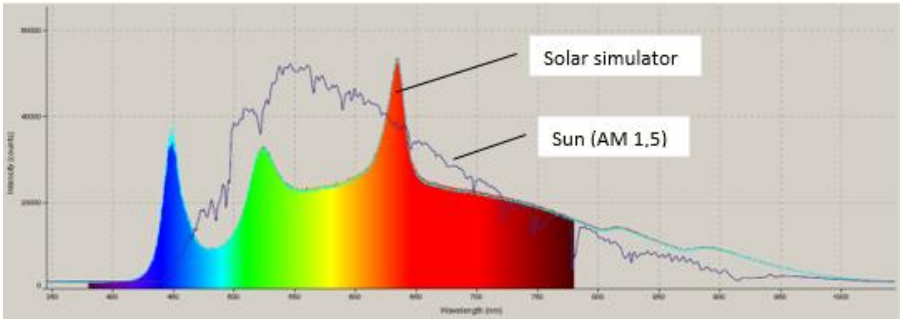

Figure 12
Spectral structure of the sun simulator and natural sunlight

Table 6
Actual spectral conformity within the inspected wavelength ranges [5-8]

Wavelength range [μm]	0.4-0.5	0.5-0.6	0.6-0.7	0.7-0.8	0.8-0.9	0.9-1.1
Spectral conformity [%]	120.9	59.0	89.3	116.8	160.6	182.7

The rate of the measured spectral conformity is always within the 40% - 200% range assigned to a class C sun simulator, hence the desired spectral structure can be provided [10, 11]. The unexpected correctness of the lowest wavelength interval is caused by a slight shift in the intensity peak of the blue LED lights, therefore the application of UV dominant LED units may not be necessary. Many other researchers received similar results, inter alia, Kenny et al. [20] and Hussain et al. [25].

4.3 Inspection of Operating Temperature

During the operation of the sun simulator, heat production caused by current running through certain units is inevitable. Excessive warming may result in the deterioration of efficiency, it can damage the device, and may result in fire- and accident hazards, hence it is important to determine if the operating temperatures of the device fall within the safe zone [15].

Three thermometer probes are placed at three points of the – initially room-temperature – device, the operating temperatures will be measured and inspected in these points. One sensor is placed atop a solar cell above the inspection range, another is placed on the surface of the heatsink below the LEDs, and the last one is on the back of the heatsink. The heatsink is cooled by a fan above it. During the measurement, the halogen lamps are operated at maximum load, the LED units are uniformly powered at 700 mA. A four-channel digital thermometer is used to measure and save the results once every second. The whole measurement is carried out in roughly 20 minutes, at which point constant temperature values are obtained at all three points [15]. The obtained results are shown in Figure 13.

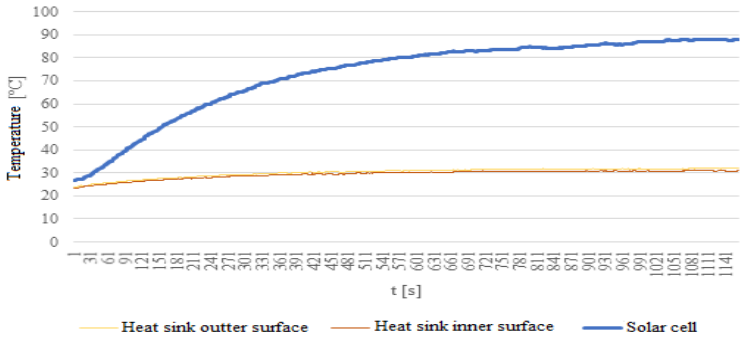

Figure 13
Temperature transient of the surface and the back of the heatsink

Even at the critical points of the device no dangerous overheating is experienced. After 20 minutes of operation, the temperature stabilizes at a constant value on both sides of the heatsink, which in both cases are around 30 °C. Measuring the temperature of the semiconductor material of the LED is not possible, thus the

dimensioning of the heatsink is not executable. However, the temperature increase on the heatsink is barely 10 °C, which is well within the generally accepted 20 °C boundary, meaning that the cooling is without doubt applicable.

As expected, the temperature of the solar cell is much higher than the temperature measured on the heatsink, since it is being illuminated by 8 halogen lamps and 36 LED units simultaneously. The temperature of the solar cell stabilized at around 88 °C. This is roughly the same temperature as it would be of a solar cell illuminated by natural sunlight on a sunny summer day. In the light of this information it can be stated that the solar cell will not suffer any damages from this amount of heat, although its temperature, according to plans, is going to be much lower. In conclusion, the sun simulator can be declared safely operable in terms of heating.

Results and Conclusions

Herein, we provide an overview of the whole work, in which, the design and construction of the standard sun simulator was executed. During the design, fixed parameters and, with the help of optimizing calculations, variable parameters were defined, while always keeping the proper light intensity distribution and the spectral structure of the illumination. Then the sun simulator was assembled, and test measurements were implemented.

The light intensity distribution was inspected, using a light intensity meter, in cases with different halogen lamp positions. We achieved an inhomogeneity of 9.96%, which is acceptable for a standard class C sun simulator. The average of the light intensity in this case is 951 W/m^2, which is slightly below the 1,000 W/m^2 peak, but is a good approximation of an average light intensity on a clear day, hence it is applicable to inspect solar cells.

The spectral composition of the illumination was inspected using a spectrometer. According to the design, proper spectral conformity was expected in each interval but one; between the wavelengths of 400 nm – 500 nm. In possession of the measurement results, it can be said that the spectral conformity of the sun simulator is within the prescribed class C standards, meaning that the methods used during the design of the sun simulator were correct and are applicable to provide close to reality calculations.

An important aspect when designing the sun simulator is to keep the costs realistic. For small sized solar cells, a standard illumination can be provided by a relatively inexpensive device; the full price of this construction is 600 Euro.

Acknowledgement

“The described article was carried out as part of the EFOP-3.6.1-16-2016-00011 “Younger and Renewing University – Innovative Knowledge City – institutional development of the University of Miskolc aiming at intelligent specialization” project implemented in the framework of the Szechenyi 2020 program. The

realization of this project is supported by the European Union, co-financed by the European Social Fund."

References

[1] Wang, W.: Simulate a 'Sun' for Solar Research: A Literature Review of Solar Simulator Technology, Royal Institute of Technology, Department of Energy Technology, Swede, Stockholm, 2014, p. 7

[2] Kohraku, S., Kurakowa, K.: A Fundamental Experiment for Discrete-Wavelenght LED Solar Simulator, Solar Energy Materials & Solar Cells 90. 2006, pp. 3364-3370

[3] Bliss, M., Betts, R., T., Gottschlag, R.: An LED-based photovoltaic measurement system with variable spectrum and flash speed, Solar Energy Materials & Solar Cells 93, 2009, pp. 825-830

[4] Kim, A, K., Dostart, N., Huynh, J., Krein, T., P.: Low-Cost Solar Simulator Design for Multi-Junction Solar Cells in Space Applications, Power and Energy Conference at Illinois, 2014, p. 6

[5] Bodnár, I.: (2018) Electric parameters determination of solar panel by numeric simulations and laboratory measurements during temperature transient, Acta Polytechnica Hungarica, Vol. 15, No. 4, 2018, pp. 59-82

[6] Kativar, M., Balkom, M., Rindt, C. C. M., Keizer, C., Zondag, H. A.: Numerical model for the thermal yield estimation of unglazed photovoltaic-thermal collectors using indoor solar simulator testing, Solar Energy, Vol. 155, No. 10, 2017, pp. 903-919

[7] Tawfik, M., Tonnellier, X., Sansom, C.: Light source selection for a solar simulator for thermal applications: A review, Renewable and Sustainable Energy Reviews, Vol. 90, 2018, pp. 802-813

[8] Kádár, P., Varga, A.: Measurement of spectral sensitivity of PV cells. 2012 IEEE 10th Jubilee International Symposium on Intelligent Systems and Informatics. 2012, pp. 549-552

[9] Chegaar, M., Ouennoughi, Z., Hoffmann, A.: A new method for evaluating illuminated solar cell parameters, Solid-State Electronic, Vol. 45, No. 2, 2001, pp. 293-296

[10] Varga, A., Libor, J., Rácz, E., Kádár, P.: A small laboratory-scale experimental method and arrangement for investigating wavelength dependent irradiation of solar cells, Proceedings of the 11th IEEE International Symposium on Applied Computational Intelligence and Informatics, 2014, pp. 137-141

[11] Munoz-Garcia, M. A., Marin, O., Alonso-García, M. C., Chenlo, F.: Characterization of thin film PV modules under standard test conditions:

Results of indoor and outdoor measuremnets and the effects of sunlinght exposure, Solar Energy, Vol. 86, No. 10, 2012, pp. 3049-3056

[12] González, M. I.: An LED solar simulator for student labs, Physics Education, Vol. 52, No. 3, 2017, p. 5

[13] Sowmy, D. S., Ara, P. J. S., Prado, R. T. A.: Uncertainties associated with solar collector efficiency test using an artificial solar simulator, Renewable Energy, Vol. 108, 2017, pp. 644-651

[14] Siddiqui, R., Kumar, R., Jha, K. G., Morampudi, M., Rajput, P., Lata, S., Agariya, S., Nanda, G., Raghava, S. S.: Comparison of different technologies for solar PV (Photovoltaic) outdoor performance using indoor accelerated aging tests for long term reliability, Energy, Vol. 107, No. 15, 2016, pp. 550-561

[15] Gu, R., Ding, J., Wang, Y., Yuan, Q., Wang, W., Lu, J.: Heat transfer and storage performance of steam methane reforming in tubular reactor with focused solar simulator, Applied Energy, Vol. 233-234, 2019, pp. 789-801

[16] Moss, R. W., Shire, G. S. F., Eames P., Henshall P. Hyde T., Arya F.: Design and commissioning of a virtual image solar simulator for testing thermal collectors, Solar Energy, Vol. 159, 2018, pp. 234-242

[17] Boubault, A., Yellowhair, J., Ho, C. K.: Design and Characterization of a 7.2 kW Solar Simulator, Journal of Solar Energy Engineering, Vol. 139, No. 3, 2017, p. 8

[18] Esen, V., Saglam, S., Oral, B.: Light sources of solar simulators for photovoltaic devices: A review, Renewable and Sustainable Energy Reviews, Vol. 77, 2017, pp. 1240-1250

[19] Wang, W., Aichmayer, L., Garrido, J., Laumert, B.: Development of a Fresnel lens based high-flux solar simulator, Solar Energy, Vol. 144, 2017, pp. 436-444

[20] Kenny, R. P., Vigan, D., Salis, E., Bardizza, G., Norton, M., Müllejans, H., Zaaiman, W.: Power rating of photovoltaic modules including validation of procedures to implement IEC 61853- 1 on solar simulators and under natural sunlight, Paper Presented at 27TH EU PVSEC 2012, Frankfurt, Germany, Photovoltaics, Vol. 21, No. 6, 2013, pp. 1384-1399

[21] Jiang, H., Lu, L., Sun, K.: Experimental investigation of the impact of airborne dust deposition on the performance of solar photovoltaic (PV) modules, Atmospheric Environment, Vol. 45, No. 25, 2011, pp. 4299-4304

[22] Chen, F., Wittkopf, S. K., Ng, P. K., Du, H.: Solar heat gain coefficient measurement of semi-transparent photovoltaic modules with indoor calorimetric hot box and solar simulator, Energy and Buildings, Vol. 53, 2012, pp. 74-84

[23] Hamadani, B. H., Chua, K., Roller, J., Bennahmias, M. J., Campbell, B., Yoon, H. W., Dougherty, B.: Towards realization of a large- area light-emitting diode- based solar simulator, Photovoltaics, Vol. 21, No. 4, 2013, pp. 779-789

[24] Guo, P., Wang, Y., Meng, Q., Li, J.: Experimental study on an indoor scale solar chimney setup in an artificial environment simulation laboratory, Applied Thermal Engineering, Vol. 107, 2016, pp. 818-826

[25] Hussain, F., Othman, M. Y. H., Yatim, B., Ruslan, H., Sopian, K., Anaur, Z., Khairuddin, S.: Fabrication and Irradiance Mapping of a Low Cost Solar Simulator for Indoor Testing of Solar Collector, Journal of Solar Energy Engineering, Vol. 133, No. 4, 2011, p. 4

[26] Precup, R. E., Preitl, S.: Stability and Sensitivity Analysis of Fuzzy Control Systems, Mechatronics Applications, Acta Polytechnica Hungarica, Vol. 3, No. 1, 2006, pp. 61-76

[27] Pozna, C., Precup, R. E., Tar, J. K., Skrjank, I., Preitl, S.: New results in modelling derived from Bayesian filtering, Knowledge-Based Systems, Vol. 23, No. 2, 2010, pp. 182-194

[28] Ürmös, A., Farkas, Z., Farkas, M., Sándor, T., Kóczy, L. T., Nemcsics, Á.: Application of self-organizing maps for technological support of droplet epitaxy, Acta Polytechnica Hungarica, Vol. 14, No. 4, 2017, pp. 207-224

[29] Vrkalovic, S., Lunca, E. C., Borlea, J. D.: Model-free sliding mode and fuzzy controllers for reverse osmosis desalination plants, International Journal of Artificial Intelligence, Vol. 16, No. 2, 2018, pp. 208-222

Optimization of an Off-Grid Hybrid System Using Lithium Ion Batteries

Paul Arévalo Cordero[1], Juan Lata García[2], Francisco Jurado[1]

[1]Department of Electrical Engineering of the University of Jaén, 23009 Jaén - Spain*; E-mails: wpac0001@red.ujaen.es, fjurado@ujaen.es

[2] Department of Electrical Engineering Salesite Polytechnic University GIPI, 090101 Guayaquil-Ecuador; E-mail: jlatag@ups.edu.ec

Abstract: Distributed generation represents an optimal solution for energy problems in remote sites, where access to energy distribution companies is difficult. This article therefore presents the modeling and optimization of an electric power system based on hybrid renewable energy (photovoltaic/hydrokinetic/diesel), combined with a Lithium Ion battery storage system. The analysis is performed using two methods. The first is a computational model based on mathematical equations using the Simulink Design Optimization (SDO) tool in Matlab. The second is a simulation of the hybrid system using the HOMER Software. The results show that the highest annual energy contribution corresponds to the hydrokinetic generation system (78.87%), the photovoltaic generation with a percentage of 19.59% and the penetration of the diesel generator is the lowest with 1.53%. In addition, the Nelder-Mead algorithm (SDO) shows a lower number of diesel generator starts, increasing the capacity of the batteries.

Keywords: Hydrokinetic turbine; Solar panels; Renewable energy; Hybrid System

1 Introduction

Considering the population increase, the reduction of oil reserves and evident negative climate change produced by a hydrocarbon based economy, renewable energies should be further developed in the coming years. Therefore, further research in this field is necessary. Furthermore, the planet can have enough renewable resources to cover global electricity demand, the energy being the main source of development, especially in countries where not the entirety of the population has electricity [1]. However, in Ecuador, there are few rural electrification projects and even less in Amazonian populations such as Kichwa, Siona and Achuar, which require electrical energy for development in their daily activities. However, in Ecuador there are few rural electrification projects and even less in Amazonian populations. Therefore, it is necessary to carry out

research projects on possible electrification taking advantage of the natural resources of rural areas [2].

Electricity service in Ecuador is 97.04%. However, the lowest rate of electrification is present in the Amazon region [1]. The deficit of electricity in these regions is the result of limited reach of distribution companies, but the main reason is the remote area of these communities, since they are located hundreds of kilometers away from the main distribution centers. For this reason, the viability of generating electricity with renewable sources in the consumption zone is analyzed. The Agency for Regulation and Control of Electricity (ARCONEL) authorized the construction of photovoltaic energy projects in 2018, with a preferential rate of USD 0.40 per kilowatt hour (kWh) of generation, since the country has a high rate of solar resource [1].

In Ecuador, one of the main sources of electricity generation is hydroelectric, since the country is located in the middle of the Andes Mountain Range, therefore it has high water resources.

According to the Ministry of Electricity and Alternative Energies, in 2023, the country will have more than 90% of hydroelectric generation. In the case of remote communities, a practical solution could be to generate electricity with hydrokinetic turbines located in the consumption sites. Unlike a hydroelectric plant, hydrokinetic turbines do not require the construction of dams for reservoirs, they simply require a continuous flow of water. There are recent studies of technical-economic feasibility for the use of this type of turbines in remote communities with favorable results [3] [4]. The "Nuevo Rocafuerte" community does not have 100% electricity coverage. Nowadays, electricity is produced by diesel genera-tors, even though it is located next to the Napo River, therefore it has a lot of water [5].

Considering the abundant water and solar resources in the area, two renewable sources could be combined in a hybrid electric power generation system (PV / HKT / DG / Li-ion). Lithium Ion batteries have a higher energy density than the acid lead, apart from having a life cycle greater than 10 years, have greater depth of discharge (SOC) [6]. In this way, the demand of the community could be efficiently supplied by hybrid systems [7]–[13]. The proposed hybrid system will be analyzed through the HOMER software [14], that allows to determine the optimal configuration of the system, and a simulation in Matlab Simulink Design Optimization (SDO), in order to ensure the best configuration and energy reliability of the community. The determination of the sizing of a hybrid system of renewable energies is increasingly notable. Different authors have investigated methods that guarantee the optimization of the system, adding multiobjective functions and studies of demand prediction, [15]. The hybrid system that is most researched is the combination of photovoltaic and wind energy. The reference [16], explains the sizing of the hybrid system composed of (PV-Wind) in a remote

area of Tunisia obtaining the results of sizing optimization, and other isolated systems [17]–[26].

In the reference [14], performs an economic feasibility analysis by using a hybrid system to reduce environmental pollution, using different fuels to find the best configuration using Homer software, showing that a hybrid system can supply the demand without violating any restrictions.

In relation with the storage system, the most common ones are lead acid batteries and hydrogen tanks. The backup systems are typically DG and electrolysers (FC), in some cases supercapacitors [27], today more authors use lithium ion batteries as a storage system due to their high efficiency [6] [28].

In reference [4], determines the optimal sizing of a Hybrid System of Renewable Energies (PV / HKT / DG) in an island located on the Ecuadorian coast. The study demonstrates that the system can supply off-grid loads, which could be analyzed on a large scale, especially evaluating the stability of the system with high penetration of renewables, however it could be further optimized with the use of Lithium Ion batteries, taking into consideration the impact on the hybrid system using this storage technology.

Renewable systems have proven to be an effective alternative by reducing the consumption of fossil fuels. The reference [29], explains the gradual reduction of diesel consumption in a remote community of Indonesia using solar panels and batteries. Moreover, in [30] analyzes energy management and optimization of battery size for a micro-grid located on an island in Australia with novel results.

On the other hand, photovoltaic and hydroelectric solar energy has been developed in Ecuador during the last decade [31]. Today, there are many studies of hybrid renewable systems in remote communities encouraging the use and optimization of renewable resources. The study conducted by [4], explains that a hybrid system (PV/HKT/DG/Lead acid Battery) has the capacity to supply the demand of an off-grid community. It is noteworthy that the acid lead battery has a short life cycle. In the same way, [25] discusses replacing the diesel generation system with fuel cells, yet fuel cells are expensive. In addition, this work may be extended in the future to include hydrogen and different storage systems.

The difficulty in designing the sizing of renewable systems focuses on the randomness of resources for their sources, in conventional systems the information is insufficient to simulate and predict complex systems. In reference [32], the author shows the use of fuzzy cognitive maps, with the aim of representing dynamic systems by means of migration algorithms, as convenient means of adaptation. Through this method it has been possible to combine several variables to approach the optimal solution of a renewable system. Research based on improving the behavior of renewable energy sources proposes different methods, as in the case of [33], this paper proposes a convergent IFT algorithm for single-entry discrete time systems using fuzzy control logic, this type of control

systems can process non-linearity is better than other control systems, although the error is non-zero, but this type of control could be useful to help the convergent algorithms that are part of renewable systems through an adequate analysis of stability and convergence. The improvement in the use of optimization algorithms are increasingly evident, in [34] has presented an ideal gas optimization algorithm (IGO), which is based on the first law of thermodynamics and kinetics, basically uses the grouping of gas molecules in each iteration to find non-uniform spaces finding suboptimal. This algorithm has a greater capacity and efficiency of analysis than some highly used algorithms, such as particle swarm (PSO) and genetic algorithm (GA). Similarly, different types of algorithms are discussed in [35]–[38], these studies present the different types of algorithms inspired by nature, focused mainly on diffuse control (Takagi-Sugeno, Swarm of Particles, Gravitational Search), hybrids (Swarm of Particles-Gravitational Search) and adaptive (Gravitational Search).

In the papers [9] [39], several optimization algorithms have been used, among them: basic equation, SDO (genetic algorithm, latyn hypercube) and Hoga and Homer software, in order to find optimal values in a hybrid system composed of PV/WT/FC systems, the results show that the best configuration found corresponds to the SDO method, performing a technical optimization through the application of genetic algorithm.

Similarly, in [12] the same hybrid system has been analyzed. However, the simulation was carried out only in Matlab software, through 3 dispatch strategies combining lead acid batteries and FC cells. The results have shown that, combining both storage systems, the results are more energy efficient. On the other hand, reference [40] has carried out an optimization analysis of a hybrid system composed of HKT/PV/lead acid batteries, using two methods: SDO and Homer software, the results have shown that the best result is found through SDO (Latin Hypercube algorithm), reducing the operation of the DG. In addition, this article has obtained similar results. The Matlab software through its SDO tool has obtained the best technical result using the Nelder-Mead algorithm, reducing the penetration of the diesel generator.

Several authors present problems of optimization with respect to hybrid systems composed of renewable energy, such as Homer software, which is highly accepted by the scientific community and has several libraries that facilitate the analysis of data. On the other hand, Matlab has many tools to improve the optimization of the parameters of a system and can be edited as necessary. Therefore, it is important to analyze a renewable hybrid system using both methods [9].

El software Matlab (SDO) puede proporcionar opciones de simulación en el dominio del tiempo y la frecuencia, utilizando la interfaz gráfica de usuario (GUI). In addition, it is possible to impose limits for the study of several signals. The design variables need to be declared in SDO as well as their limits, for example the SOC of a lithium ion battery, the charge limits of the battery must be

maintained in SOCmin of 20% and SOCmax of 100%. Then, the variables to be optimized must be defined, such as the size of the batteries, PV and HKT. In addition, it is possible to choose the optimization method in SDO, among them: gradient gescent, pattern search and simplex search. In several cases, it has been concluded that the pattern search method is able to find the appropriate parameters in the simulation. Within this method, optimization algorithms such as: Latin hypercube, genetic algorithm and Nelder Mead, should be chosen. Finally, the following parameters are defined for the optimization of the control process [9].

Parameter tolerance: the optimization process ends when successive parameters values change less than this number.

Constraint tolerance: this parameter represents the maximum relative amount by which the constraints can be violated and still allow a successful convergence.

Function tolerance: this parameter is used to end the optimization when successive function values are less than this value. This parameter is only useful when Simplex Search method is selected. Its value does not matter in our study.

Maximum iterations: define the maximum number of iteration allowed in the optimization process.

The general objective of this study is to optimize hybrid system using Lithium Ion batteries in order to analyze the resulting effects. The originality of this study is reflected in the type of storage system, which by the combination of two renewable sources must provide clean and reliable energy to the community under study. In addition, the battery replacement time should be reduced. Therefore, this study is organized as follows: mainly, the study area is analyzed, obtaining data such as solar radiation, river velocity and electric demand, then the mathematical modeling of the generation sources (PV/HKT/DG) and storage of Lithium Ion (Li-ion), to later perform the simulation using the Matlab tool (Simulink design optimization) and in the Homer software. Hence, by obtaining system sizing results that will be analyzed with the purpose of determining the optimal configuration.

2 Case Study

The case study is the "Nuevo Rocafuerte" community, part of the Aguarico Canton located in the province of Orellana in the Ecuadorian Amazon, see Figure 1. This community is located at a latitude of 0 ° 56'00 " South and a length of 75 ° 24'00 " West. The number of inhabitants until the last census conducted in 2010 was 1,024, according to the National Institute of Statistics and Census (INEC). The economy is mostly based on agriculture and fishing.

Today, according to the National Electricity Corporation (CNEL) in the area of "Nuevo Rocafuerte" there are electric generators powered by diesel, which operate from 09:00 to 24:00, making it difficult for most of the inhabitants to carry out their activities outside of the established schedule. On the other hand, building electricity transmission lines is environmentally and economically unfeasible, since the community is surrounded by two natural reserves.

Figure 1
Community "Nuevo Rocafuerte", province of Orellana
Source: (National Institute of Statistics and Census - INEC)

2.1 Demographic Profile and Meteorological Data of the Studied Community

The inhabitants of the area cannot enjoy the comforts provided by a power generation system that supplies electricity all day. The annual average solar radiation in the area is 4.35 kWh/m^2/day and the clarity index is 0.43 (NASA Prediction of Worldwide Energy Resources). Figure 2 represents the hourly solar radiation of the selected area using the Homer software database [41].

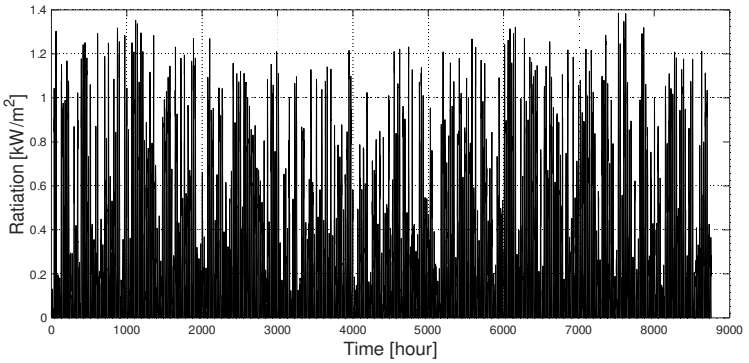

Figure 2
Solar Radiation in the community

Additionally, the flow (m^3/ s) and the monthly average speed (m/s) of the Napo River are shown in Table 1, with a value of 2.34 m/s.

Figure 3 presents the speed of the river as developed by the HOMER software. It is demonstrated that there is a variation range of 1.8 m/s to 4 m/s throughout the year, with the months of October, November and December being those with the greatest variations. This information will be considered later in the analysis and optimal sizing of the hydrokinetic turbine.

Table 1
Contains the Flow and speed of the Napo River in "Nuevo Rocafuerte"

Flow (m3/s)	Months	River speed (m/s)
1500	Jan	2.591
1600	Feb	2.656
1800	Mar	2.526
2100	Apr	2.62
2900	May	2.859
3000	Jun	2.81
2600	Jul	2.739
2200	Aug	2.46
1800	Sep	2.591
1800	Oct	2.227
1800	Nov	2.469
2100	Dec	2.54

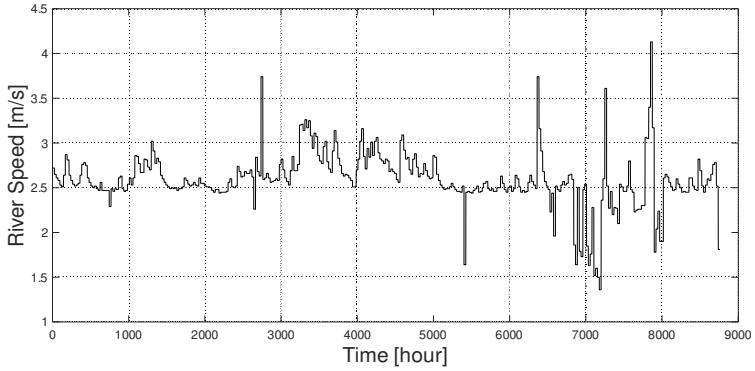

Figure 3
River Speed

2.2 Estimation of Electricity Consumption Community

The habits of the inhabitants of the area under study do not vary substantially with respect to the rest of Ecuadorians. According to the Electricity Corporation of

Ecuador (CELEC EP), the maximum of the national electricity demand occurs between 19:00 to 20:00, and the minimum between 01:00 and 05:00. Since there is no exact data of the demand of the inhabitants of the area, an approximation is made regarding the national demand (See Figure 4).

When it comes to energy consumption, a group of loads is considered typical of a household located in rural areas, it is assumed the use of LED lighting and efficient refrigerators of reduced size. Table 2 indicates the common alternating current loads for the system. It should be mentioned that DC loads are not analyzed, since they are not common, even for this type of applications. The loads presented are estimated according to the customs and climate of the area.

Table 2
Customers Load

Description	Quantity	Current (A)	Power (W)	Hours / day	Days / week	Weekly consumption (Wh)
LED light	3	0.1	10	4	7	280
TV	1	0.2	25	6	7	1,050
Radio	1	0.15	20	3	7	420
Cell charger	1	0.05	5	5	3	75
Alarm clock	1	0.05	5	24	7	840
Installed power (W)	Simultaneous power (W)	Weekly energy (Wh)	Losses	Monthly energy (Wh)		
65	52	2,665	0.05	10,127		

According to Table 2, the demand in each house is 65 W on average, since the population of "Nuevo Rocafuerte" is approximately 1,000 inhabitants. Considering the average family consisting of 3 to 4 members including street lighting, the total installed power would be approximately 20 kW. It can be seen that peak demand occurs between 19:00 and 20:00.

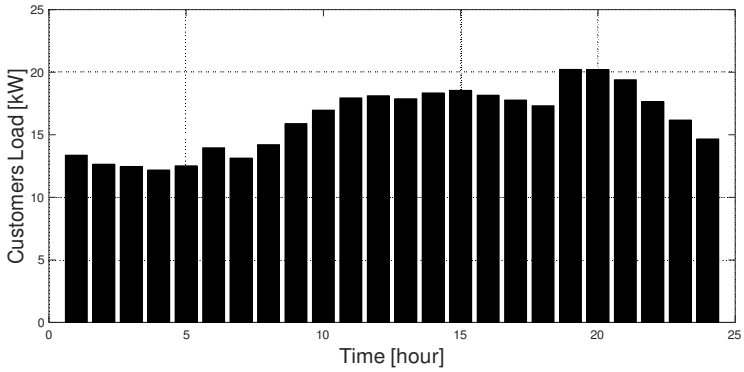

Figure 4
Daily load profile

There are no significant variations during the year (Figure 5).

Figure 5 presents the seasonal load profile obtained by [19], based on real data from the area. To model the system, it is very important to use one hour intervals during a year.

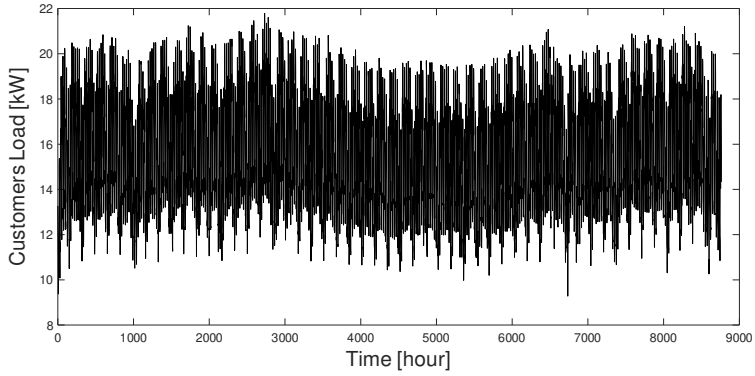

Figure 5
Seasonal load profile

3 Modeling - Sources of Generation and Storage

3.1 Modeling of the Photovoltaic Solar System (PV)

The performance of a solar module varies according to the climatic conditions of the installation site [42].

The photovoltaic power varies according to the incident solar radiation in the module and the environmental factors (ambient temperature, humidity, etc.). The system power is calculated using equation (1).WhereY_{PV} (kW) is the photovoltaic generator nominal capacity, f_{PV} (%) is the reduction factor I_T (kW/m^2) is the solar energy from the incident irradiation in the photovoltaic module, I_S (kW/m^2) is the incident solar radiation under standard test conditions, α_P is the temperature coefficient of the power, T_c (˚C) is the PV cell temperature, and T_s (°C) Temperature of the photovoltaic cell under standard test conditions (25 ° C) [42].

$$P_{PV} = Y_{PV} f_{PV} \cdot \left(\frac{I_T}{I_S}\right).[1 + a_P \left(T_c - T_s\right) \quad (1)$$

The cell temperature is obtained by equation (2).

$$T_c = T_a + I_T \frac{T_{c,NOCT} - T_{a,NOCT}}{I_{T,NOCT}} \left(1 - \frac{h_{PV}}{t\,a}\right) \quad (2)$$

Where:

T_a, Is the ambient temperature, $\tau\alpha$ is the absorption transmittance of the photovoltaic module. In general, HOMER takes it as 0.9, NOCT is the nominal temperature of the operating cell that results in a solar radiation of 800 W/m^2, at an ambient temperature of 20 °C and in unloaded conditions and η_{PV} is the panel efficiency. In the study area there is an average temperature of approximately 25 °C, the best option is to use polycrystalline solar panels [42]. This study uses the parameters of the US commercial solar module SunPower, model E20-327, 360 W of peak power with an efficiency of 17%, polycrystalline and with a power coefficient of -0.37% / °C. In this system, the tracking system and the 80% rating reduction factor have not been taken into account.

The system mode in SDO is shown below:

The function f(u) is represented by Equation (1)

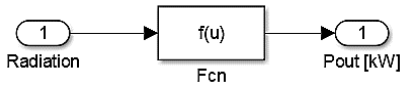

Figure 6
Simulated photovoltaic system in SDO

3.2 Modeling Hydrokinetic System (HKT)

In this study, taken into consideration the river speed characteristic, the hydrokinetic turbine has been selected (Smart Hydro Power, Feldang, Germany 5 kW). The power of the hydrokinetic turbine is given by equation (3) [25].

$$P_{HKT} = \frac{1}{2}.r\,w.A.v^3.C_{p,H}.h_{HKT} \quad (3)$$

The energy is given by:

$$E_{HKT} = P_{HKT}.t \quad (4)$$

It is important to consider that the river velocity profile applies to open channel.

In the open channel flow, the velocity of the river at a certain point above the river bed is considered a power function that uses the depth relationship. The average velocity in time (u), the normal distance of the rising bed above the reference (y), the velocity at the point ''a'' where it deviates vertically from the river bottom (u_a) and the exponent of the power law $(1/m)$, [4].

$$\frac{u}{u_a} = \left(\frac{y}{y_a}\right)^{\frac{1}{m}} \tag{5}$$

The system mode in SDO is shown below:

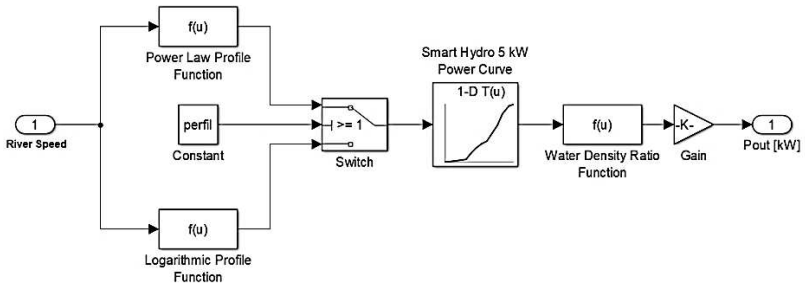

Figure 7
Hydrokinetic turbine in SDO

3.3 Modeling of the Diesel Generator (DG)

As mentioned in the previous section, the supply of electrical energy from renewable sources is random depending on the available resource. To ensure the reliability of the supply, it has been decided to use diesel generators as a backup system in case there is no solar radiation in times of low water and if the bank of batteries is discharged. HOMER calculates the capacity of the diesel generator with an additional 10% to the highest peak of present demand with a life expectancy of 15000 h. The maximum load ratio of the generator is 25%. The fuel consumption for the production of electricity is calculated according to equation (6) [43].

$$F = F_{0,dg} Y_{dg} + F_{1dg} P_{dg} \tag{6}$$

Where, F_0 is the intercept coefficient of the fuel curve ($0.000205 \mathrm{m}^3/\mathrm{h}$), F_1 is the slope of the fuel curve ($0.00025\ \mathrm{m}^3/\mathrm{h/kW}$), Y_{dg} (kW) is the rated capacity of the generator and P_{dg} (kW) is the electrical power. The lowest calorific value of diesel fuel is $43.2\ \mathrm{MJ/kg}$ with a density of $820\ \mathrm{kg/m}^3$ [43].

The work factor (DF) (kWh/start-stop/year) is the power generation ratio of the supplementary primary motors to the total start-stop and can be calculated using equation (7), where $N_{s/s}$is the number of starts and stops [44].

$$DF = \frac{P_{dg}}{N_{s/s}} \tag{7}$$

The energy dispatch of the diesel generator will be carried out depending on the hybrid system conditions. When the generation and demand are disproportionate, a shutdown of the electrical system can be produced since its generation sources would not be able to maintain the load.

The generation, (PV + HKT) must be equal to the consumption (Load + Pbat). When this equality condition is not met, the diesel generator must be turned on to supply the demand. Also, if the objective function (PV + HKT-Load-Pbat) is in the stability zone, that is, the generation must be greater than or equal to the demand, the generator must be turned off. For the simulation a first-order model is used so that the fuel consumption is a function of the mechanical power and speed of the engine [45].

The system mode in SDO is shown below:

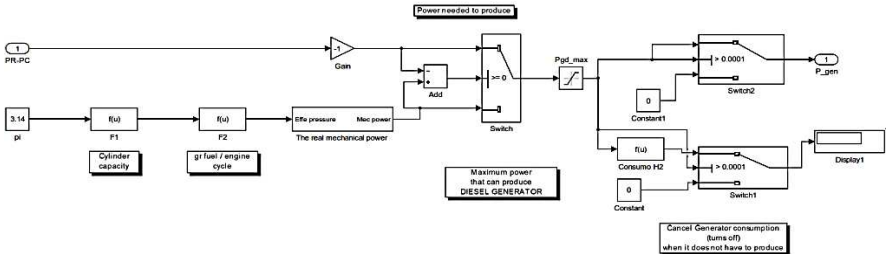

Figure 8
Diesel generator in SDO

3.4 Modeling of the Storage System (Li-ion)

In this study, a Lithium Ion battery bank is used as a storage system. The process of charging the batteries begins when the generation of electric power is greater than the consumption. In the same way, when the renewable system is not sufficiently capable of producing energy according to the demand (dry season or low solar radiation), the load will be supply through the storage system. Considering the state of charge (SOC) of the batteries, a charge and discharge limit can be established.

The capacity of a battery is defined as the maximum charge that can be extracted from a battery when it is fully charged. [46], (International Standard IEC 61427-61432). According to [19], several equations are used to determine the charge and discharge power of the Lithium Ion battery. Equation (8) determines the maximum load power of the storage system (battery).

$$P_b(t) = \frac{kQ_1(t)e^{-k} + Q(t)kc\left(1 - e^{-kDt}\right)}{1 - e^{-kDt} + c\left(kDt - 1 + e^{-kDt}\right)} \tag{8}$$

Where: $Q_1(t)$refers to the energy available at the beginning of the operating range and above the minimum load state (SOC), in the case of Lithium Ion batteries is 20%, the maximum SOC will be 100%. Q (t) it is the total energy at the beginning of the passage of time, c is the ratio of the storage capacity of each system, k is the constant energy storage rate, and Δt is the time interval. The maximum power can be calculated with the following equation:

$$P_b(t) = \frac{kQ_{max}(t)e^{-k} + kQ_1(t)e^{-k} + Q(t)kc\left(1 - e^{-kDt}\right)}{1 - e^{-kDt} + c\left(kDt - 1 + e^{-kDt}\right)} \tag{9}$$

Where $Q_{max}(t)$ is the total storage capacity.

Homer calculates the life span of the battery according to its performance. In this case different parameters that help to choose the optimal combination of the system are compared.

In order to determine the number of batteries necessary, factors such as the life span of the hybrid system and that of every battery should be considered [19].

$$N_{bat} = ceil.\frac{(Vida_{SH}.Vida_{bt}^{pu,año})}{\left(T_{bat}^{vida}\right)} \tag{10}$$

Where:

N_{bat} - batteries number

T_{bat}^{vida} - time period from beginning of the year to the last battery is replaced

$Vida_{SH}$ - life cycle of the hybrid system (25 years in this case)

$Vida_{bt}^{pu,año}$ - battery life in the last year

Ceil - sum of the battery bank

The system mode in SDO is shown below:

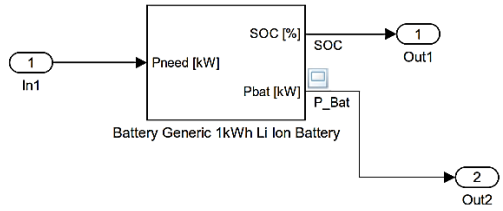

Figure 9
Storage system control in SDO

4 Energy System Control

Starting from the objective function (PFV + HKT-Load-Pbat), the power of the PV, HKT and the storage system composed of Lithium Ion batteries (Pbat), are defined as adjustment variables. (DG) is set as a fixed variable, given the fact that the continuity of the electricity supply must be ensured under any contingency.

The limit values of each parameter depend on the expressed mathematical model. For example, the battery state of charge will indicate the operating time of the diesel generator according to the load and weather conditions at each instant. Whereas, the work objective is to reduce the diesel generator operation hours without removing the continuity of the electric service, the state of charge of the batteries is defined as an input signal.

In normal climatic conditions, the demand will be satisfied by the (PV/HKT), with fully charged batteries, in times of high demand between 19:00 and 21:00, there will only be HKT generation. If there is a generation deficit, the storage system provides the energy to meet the demand. In case of hydrokinetic power deficit, the battery bank must supply all the available electricity at night time (PV power =0). When the batteries are discharged at a minimum SOC level (20%), the diesel generator should start charging the batteries as a priority. The generator must be turned off when the renewable power is greater than the demand and the batteries are charged.

In this study, the option of optimization was selected through Pattern Search in Matlab, since it allows finding appropriate values for the adjustment parameters. In addition, three of the five algorithms available for the search of patterns are considered, in order to ensure that the convergence is correct. It has been selected within the toolbox “Matlab/Simulink Design Optimization”, the Genetic Algorithm, Latin Hypercube and the Nelder-Mead, since they obtain values that correspond to those obtained by using HOMER, as indicated in Figure 10.

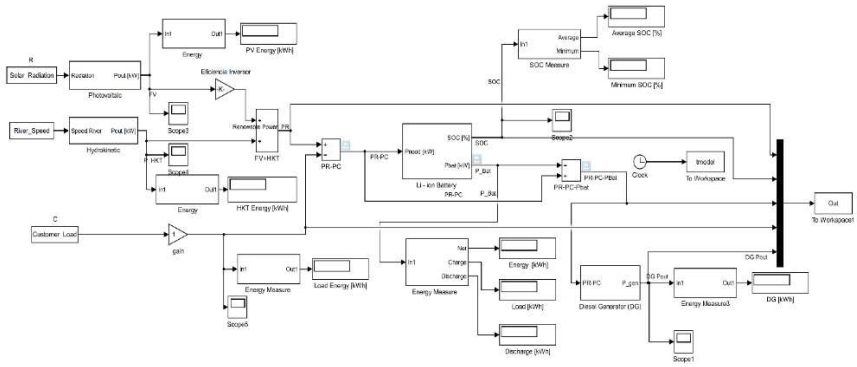

Figure 10
Simulation in Simulink Design Optimization

Figure 11 depicts the simulated Homer system, where both the batteries and the photovoltaic system are connected to a direct current bus. The (HKT, DG and Customers load) are connected to the alternating current bus. The energy is taken from the PV/Li-ion system to the load though an inverter, and the excess energy is stored in the batteries by means of a rectifier (Converter).

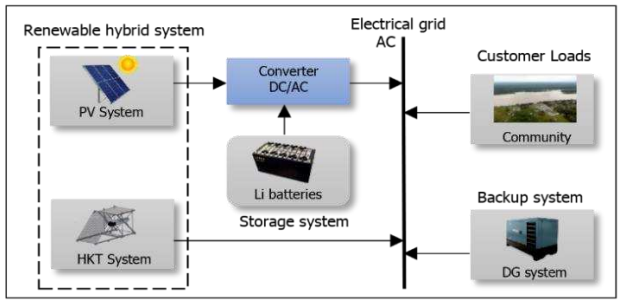

Figure 11
Schematic of the Hybrid System

5 Results and Comments

Figure 12 shows the operating logic for the proposed system. It is evident that, this type of control allows the diesel generator to start only if the batteries need to be recharged and not directly to feed the electrical demand, as explained in Section 4. This insures a reduction of the DG operational hours. In Fig. 8, this type of system behavior is observed. If the Customers Load (kW) is greater than Renewable Pout (kW) and when the State Of Charge (%) is less than SOCmin (20%), the diesel generator starts at full load recharging the batteries, until the Li-ion have been

charged at least 80% in such a way as to supply the load, at that point the diesel generator turns off.

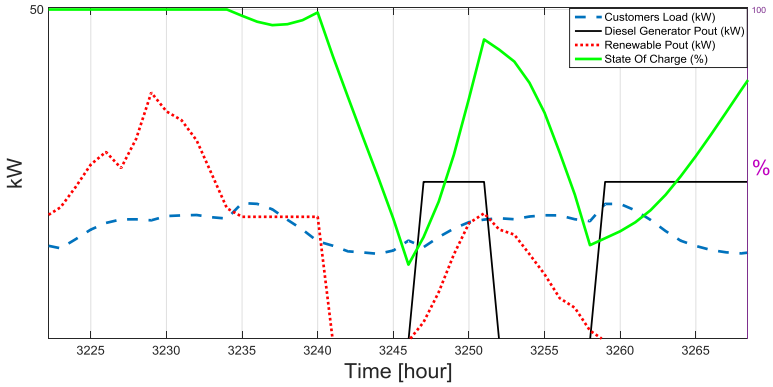

Figure 12

System operation

The results of the energy production necessary to supply the annual demand are shown in Figure 13. These results have been obtained through HOMER.

The largest contribution corresponds to the hydrokinetic generation system, with 78.87%. The photovoltaic generation represents 19.59% of the total energy production and finally the contribution of the diesel generator is merely 1.53%.

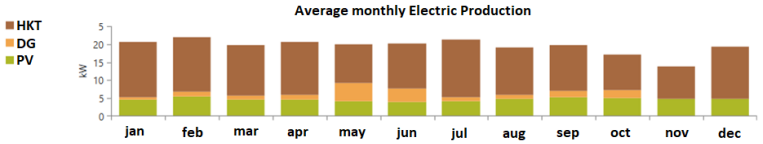

Figure 13

Electric Production

Figure 14 shows SOC of the storage system, during a year. Since October, the flow rate of river Napo is low and its velocity decreases. In this way, the batteries have to supply the load at night, since the photovoltaic system is not in operation. Here, there is an optimal system response. The generator does not allow the state of charge of the batteries to fall below 20%, thus increasing the life span of the storage system. By comparison, during the months of June and July there are enough water and solar resources to supply the demand. Figure 14 shows that in these months the batteries' SOC remains at a value close to 100%, apart from some discharge peaks that can be attributed to the known random behavior of renewable sources.

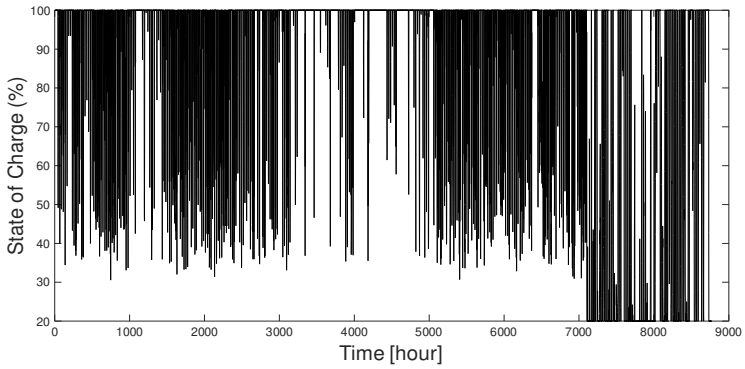

Figure 14
Li-ion batteries SOC

Conversely, the results using SDO algorithms, with respect to, Homer, are similar. Nevertheless, the contribution of the diesel generator is lower according to SDO algorithms, due to the fact that in this software, the generator is intended to operate as little as possible without taking into account extra costs in replacement of the storage system (as HOMER does).

In Tables 3 and 4 the results of the optimization of the power can be compared with Matlab and HOMER.

Table 3
Sizing and operation of the diesel generator

Algorithm	PV (kW)	HKT (kW)	DG (kW)	Li-ion (kWh)	Switch on DG/yr
Latin Hypercube	32.03	21.04	25	518.95	124
Genetic algorithm	31.05	20.56	25	513.54	155
Nelder-mead	30	20	25	530	118
Homer	30	20	24	100	237

Table 4
Annual energy produced by each component

Algorithm	PV (kWh/yr)	HKT (kWh/yr)	DG (kWh/yr)	Li-ion (kWh/yr)
Latin Hypercube	43,952	114,758	15,856	37,840.15
Genetic algorithm	43,887	114,568	14,589	32,945
Nelder-mead	43,576	114,327	11,968	48,280
Homer	41,101	115,080	17,228	3,023.9

In Table 3, the Nelder-mead algorithm calculates 118 diesel generator starts, genetic algorithm 118 starts, 124 for Latin hyper-cube and 237 for homer. However, the largest storage capacity is calculated by Nelder-mead (530 kWh). The generation sources power is similar between SDO and Homer, however, Homer calculates the lowest capacity of all the components of the system.

The energy contribution of the diesel generator is lower in Nelder-mead and the highest in the Homer result. Yet, the highest battery system result is attributed to the Nelder-mead, whereas the lowest to Homer, maintaining similar values in terms of generation by PV/HKT.

The similarity of the size of the system shown by the two simulations is evident. It is obvious that since there is a high water resource, the penetration of hydrokinetic energy is greater than the photovoltaic. The novelty is that SDO considers it viable to increase the storage system reducing the DG operation, unlike Homer, which considers it a better option to use the diesel generator more times with lower power in each one, reducing the capacity of the batteries, creating an economic optimization.

Conclusions

The results show that the hybrid system is able to supply the demand of the community throughout an entire year, since the meteorological conditions of Nuevo Rocafuerte are adequate for this purpose. With the obtained results, the use of alternative energies in Ecuador can be enhanced. In addition, there are numerous protected areas of the Amazon where rural populations are located.

The result of the research shows that it is possible to include a hybrid renewable energies system in the area, apart from the existing one that consists solely of diesel generators. One of the most important aspects shows that continuity of an electricity supply for the community is not compromised. Furthermore, his reduction of transport expenses and diesel consumption is evident, even without carrying out a cost analysis, not to mention, reducing emissions of greenhouse gases produced by DG.

The analysis carried out by Simulink Design Optimization does not take into account the minimization of system costs, since it is focused on reducing the operation of the diesel generator ensuring the continuity of the electric generation by optimizing the photovoltaic and hydrokinetic design. The variability of renewable energy sources could hinder the normal operation of hybrid system, therefore, it is recommended to perform a stability study of this system. Finally, another important aspect, is to consider is the use of Lithium Ion batteries, because in the near future, they will be used more frequently than lead based batteries. Future work will consider the use of storage systems composed of different technologies, with the aim of studying the effects they would produce within the proposed hybrid system.

References

[1] M. Ten Palomares and A. Boni Aristizabal, "Visiones de la electrificación rural en la Amazonía Ecuatoriana: disputando lógicas hegemónicas," *Let. Verdes. Rev. Latinoam. Estud. Socioambientales*, No. 20, p. 4, Oct. 2016

[2] R. Luna-Rubio, M. Trejo-Perea, D. Vargas-Vázquez, and G. J. Ríos-Moreno, "Optimal sizing of renewable hybrids energy systems: A review of methodologies," *Sol. Energy*, Vol. 86, No. 4, pp. 1077-1088, Apr. 2012

[3] H. J. Vermaak, K. Kusakana, and S. P. Koko, "Status of micro-hydrokinetic river technology in rural applications: A review of literature," *Renewable and Sustainable Energy Reviews*, Vol. 29, pp. 625-633, Jan-2014

[4] J. Lata-García, F. Jurado, L. M. Fernández-Ramírez, and H. Sánchez-Sainz, "Optimal hydrokinetic turbine location and techno-economic analysis of a hybrid system based on photovoltaic/hydrokinetic/hydrogen/battery," *Energy*, Vol. 159, pp. 611-620, Sep. 2018

[5] R. J. Cevallos Meza and M. E. Ruales Lara, "Estudio de la geodinámica actual del río Napo (Amazonía Ecuatoriana - Peruana) usando datos hidrológicos y de DGPS," 2005

[6] B. K. Das, Y. M. Al-Abdeli, and M. Woolridge, "Effects of battery technology and load scalability on stand-alone PV/ICE hybrid micro-grid system performance," *Energy*, Vol. 168, pp. 57-69, Feb. 2019

[7] C. Ammari, M. Hamouda, and S. Makhloufi, "Comparison Between Three Hybrid System PV/Wind Turbine/Diesel Generator/Battery Using HOMER PRO Software," 2019, pp. 227-237

[8] S. Ahmadi and S. Abdi, "Application of the Hybrid Big Bang–Big Crunch algorithm for optimal sizing of a stand-alone hybrid PV/wind/battery system," *Sol. Energy*, Vol. 134, pp. 366-374, Sep. 2016

[9] A. Cano, F. Jurado, H. Sánchez, L. M. Fernández, and M. Castañeda, "Optimal sizing of stand-alone hybrid systems based on PV/WT/FC by using several methodologies," *J. Energy Inst.*, Vol. 87, No. 4, pp. 330-340, Nov. 2014

[10] R. Dufo-López *et al.*, "Multi-objective optimization minimizing cost and life cycle emissions of stand-alone PV-wind-diesel systems with batteries storage," *Appl. Energy*, Vol. 88, No. 11, pp. 4033-4041, Nov. 2011

[11] M. Jamshidi and A. Askarzadeh, "Techno-economic analysis and size optimization of an off-grid hybrid photovoltaic, fuel cell and diesel generator system," *Sustain. Cities Soc.*, Vol. 44, pp. 310-320, Jan. 2019

[12] M. Castañeda, A. Cano, F. Jurado, H. Sánchez, and L. M. Fernández, "Sizing optimization, dynamic modeling and energy management strategies of a stand-alone PV/hydrogen/battery-based hybrid system," *Int. J. Hydrogen Energy*, Vol. 38, No. 10, pp. 3830-3845, Apr. 2013

[13] D. N. Luta and A. K. Raji, "Optimal sizing of hybrid fuel cell-supercapacitor storage system for off-grid renewable applications," *Energy*, Vol. 166, pp. 530-540, Jan. 2019

[14] F. K. Abo-Elyousr and A. Elnozahy, "Bi-objective economic feasibility of hybrid micro-grid systems with multiple fuel options for islanded areas in Egypt," *Renew. Energy*, Vol. 128, pp. 37-56, Dec. 2018

[15] O. Djelailia, M. S. Kelaiaia, H. Labar, S. Necaibia, and F. Merad, "Energy hybridization photovoltaic/diesel generator/pump storage hydroelectric management based on online optimal fuel consumption per kWh," *Sustain. Cities Soc.*, Vol. 44, pp. 1-15, Jan. 2019

[16] M. Belouda, M. Hajjaji, H. Sliti, and A. Mami, "Bi-objective optimization of a standalone hybrid PV–Wind–battery system generation in a remote area in Tunisia," *Sustain. Energy, Grids Networks*, Vol. 16, pp. 315-326, Dec. 2018

[17] R. Sarrias-Mena, L. M. Fernández-Ramírez, C. A. García-Vázquez, and F. Jurado, "Fuzzy logic based power management strategy of a multi-MW doubly-fed induction generator wind turbine with battery and ultracapacitor," *Energy*, Vol. 70, pp. 561-576, Jun. 2014

[18] R. Sarrias-Mena, L. M. Fernández-Ramírez, C. A. García-Vázquez, and F. Jurado, "Improving grid integration of wind turbines by using secondary batteries," *Renew. Sustain. Energy Rev.*, Vol. 34, pp. 194-207, Jun. 2014

[19] J. P. Torreglosa, P. García-Triviño, L. M. Fernández-Ramirez, and F. Jurado, "Control based on techno-economic optimization of renewable hybrid energy system for stand-alone applications," *Expert Syst. Appl.*, Vol. 51, pp. 59-75, Jun. 2016

[20] J. P. Torreglosa, P. García, L. M. Fernández, and F. Jurado, "Energy dispatching based on predictive controller of an off-grid wind turbine/photovoltaic/hydrogen/battery hybrid system," *Renew. Energy*, Vol. 74, pp. 326-336, Feb. 2015

[21] W. Zhang, A. Maleki, M. A. Rosen, and J. Liu, "Optimization with a simulated annealing algorithm of a hybrid system for renewable energy including battery and hydrogen storage," *Energy*, Vol. 163, pp. 191-207, Nov. 2018

[22] W. Zhang, A. Maleki, M. A. Rosen, and J. Liu, "Sizing a stand-alone solar-wind-hydrogen energy system using weather forecasting and a hybrid search optimization algorithm," *Energy Convers. Manag.*, Vol. 180, pp. 609-621, Jan. 2019

[23] S. C. Bhattacharyya, "Energy access programmes and sustainable development: A critical review and analysis," *Energy Sustain. Dev.*, Vol. 16, No. 3, pp. 260-271, Sep. 2012

[24] B. K. Das and F. Zaman, "Performance analysis of a PV/Diesel hybrid

system for a remote area in Bangladesh: Effects of dispatch strategies, batteries, and generator selection," *Energy*, Vol. 169, pp. 263-276, Feb. 2019

[25] J. Lata-Garcia, F. Jurado-Melguizo, H. Sanchez-Sainz, C. Reyes-Lopez, and L. Fernandez-Ramirez, "Optimal sizing hydrokinetic-photovoltaic system for electricity generation in a protected wildlife area of Ecuador," *Turkish J. Electr. Eng. Comput. Sci.*, Vol. 26, No. 2, pp. 1103-1114, 2018

[26] P. García, J. P. Torreglosa, L. M. Fernández, and F. Jurado, "Optimal energy management system for stand-alone wind turbine/photovoltaic/ hydrogen/battery hybrid system with supervisory control based on fuzzy logic," *Int. J. Hydrogen Energy*, Vol. 38, No. 33, pp. 14146-14158, Nov. 2013

[27] A. Abdelkader, A. Rabeh, D. Mohamed Ali, and J. Mohamed, "Multi-objective genetic algorithm based sizing optimization of a stand-alone wind/PV power supply system with enhanced battery/supercapacitor hybrid energy storage," *Energy*, Vol. 163, pp. 351-363, Nov. 2018

[28] A. Farmann, W. Waag, A. Marongiu, and D. U. Sauer, "Critical review of on-board capacity estimation techniques for lithium-ion batteries in electric and hybrid electric vehicles," *J. Power Sources*, Vol. 281, pp. 114-130, May 2015

[29] C. D. Rodríguez-Gallegos, O. Gandhi, M. Bieri, T. Reindl, and S. K. Panda, "A diesel replacement strategy for off-grid systems based on progressive introduction of PV and batteries: An Indonesian case study," *Appl. Energy*, Vol. 229, pp. 1218-1232, Nov. 2018

[30] K. S. El-Bidairi, H. Duc Nguyen, S. D. G. Jayasinghe, T. S. Mahmoud, and I. Penesis, "A hybrid energy management and battery size optimization for standalone microgrids: A case study for Flinders Island, Australia," *Energy Convers. Manag.*, Vol. 175, pp. 192-212, Nov. 2018

[31] M. Vasquez C., F., F, L., Espinoza, J., & Garcia R., "(12) (PDF) Energia solar en el Ecuador," 2015

[32] J. Vaščák, "Adaptation of fuzzy cognitive maps by migration algorithms," *Kybernetes*, Vol. 41, No. 3/4, pp. 429-443, Apr. 2012

[33] R.-E. Precup, M.-B. Rădac, M. L. Tomescu, E. M. Petriu, and S. Preitl, "Stable and convergent iterative feedback tuning of fuzzy controllers for discrete-time SISO systems," *Expert Syst. Appl.*, Vol. 40, No. 1, pp. 188-199, Jan. 2013

[34] M. Impr. Bourg-offset), E. Rashedi, S. M. Dashti, and A. Hakimi, *Communication, BEP terminale professionnelle : corrigé.*, Vol. 15, No. 2, Fontaine Picard, 1997

[35] R.-E. Precup, R.-C. David, R.-E. Precup, and R.-C. David, "Introduction," *Nature-inspired Optim. Algorithms Fuzzy Control. Servo Syst.*, pp. 1-54, Jan. 2019

[36] R.-E. Precup, R.-C. David, R.-E. Precup, and R.-C. David, "Nature-inspired algorithms for the optimal tuning of fuzzy controllers," *Nature-inspired Optim. Algorithms Fuzzy Control. Servo Syst.*, pp. 55-80, Jan. 2019

[37] R.-E. Precup, R.-C. David, R.-E. Precup, and R.-C. David, "Adaptive nature-inspired algorithms for the optimal tuning of fuzzy controllers," *Nature-inspired Optim. Algorithms Fuzzy Control. Servo Syst.*, pp. 81-101, Jan. 2019

[38] R.-E. Precup, R.-C. David, R.-E. Precup, and R.-C. David, "Hybrid nature-inspired algorithms for the optimal tuning of fuzzy controllers," *Nature-inspired Optim. Algorithms Fuzzy Control. Servo Syst.*, pp. 103-114, Jan. 2019

[39] M. Castaneda, L. M. Fernandez, H. Sanchez, A. Cano, and F. Jurado, "Sizing methods for stand-alone hybrid systems based on renewable energies and hydrogen," in *2012 16th IEEE Mediterranean Electrotechnical Conference*, 2012, pp. 832-835

[40] J. Lata-Garcia, C. Reyes-Lopez, F. Jurado, L. M. Fernandez-Ramirez, and H. Sanchez, "Sizing optimization of a small hydro/photovoltaic hybrid system for electricity generation in Santay Island, Ecuador by two methods," in *2017 CHILEAN Conference on Electrical, Electronics Engineering, Information and Communication Technologies (CHILECON)*, 2017, pp. 1-6

[41] J. W. White, G. Hoogenboom, P. W. Stackhouse, and J. M. Hoell, "Evaluation of NASA satellite- and assimilation model-derived long-term daily temperature data over the continental US," *Agric. For. Meteorol.*, Vol. 148, No. 10, pp. 1574-1584, Sep. 2008

[42] S. Mandal, B. K. Das, and N. Hoque, "Optimum sizing of a stand-alone hybrid energy system for rural electrification in Bangladesh," *J. Clean. Prod.*, Vol. 200, pp. 12-27, Nov. 2018

[43] F. Khalid, I. Dincer, and M. A. Rosen, "Thermoeconomic analysis of a solar-biomass integrated multigeneration system for a community," *Appl. Therm. Eng.*, Vol. 120, pp. 645-653, Jun. 2017

[44] A. Askarzadeh, "Distribution generation by photovoltaic and diesel generator systems: Energy management and size optimization by a new approach for a stand-alone application," *Energy*, Vol. 122, pp. 542-551, Mar. 2017

[45] S. Roy, O. P. Malik, and G. S. Hope, "Adaptive control of speed and equivalence ratio dynamics of a diesel driven power-plant," *IEEE Trans. Energy Convers.*, Vol. 8, No. 1, pp. 13-19, Mar. 1993

[46] M. Mehrabankhomartash, M. Rayati, A. Sheikhi, and A. M. Ranjbar, "Practical battery size optimization of a PV system by considering individual customer damage function," *Renew. Sustain. Energy Rev.*, Vol. 67, pp. 36-50, Jan. 2017

Voting to the Link: a Static Network Formation Model

Róbert Pethes and Levente Kovács

Physiological Controls Research Center
Research and Innovation Center of Óbuda University
Bécsi út 96/b, H-1034 Budapest, Hungary
pethes.robert@phd.uni-obuda.hu, kovacs.levente@nik.uni-obuda.hu

Abstract: It is well known that the structure of social, organization and economic networks have a huge effect on the behaviour of the underlying system. This structure is often considered as a network, and modelling the formation of these networks is an active research area of complex systems. In this paper we present a simple network generation model: there is a closed population of agents, and the agents are voting to the connections according to their fixed preferences. These preferences are denoted by real numbers, and can be considered as simply the desired number of connections or social capital.

Keywords: static random graph model;inhomogeneous random graphs; social networks

1 Introduction

Networks play a central role in determining the outcome of many social and economic relationships as well in the effectiveness of organizations such as companies. The fact, that we live in hierarchies [1] is a fundamental feature of our society and one of the basic organization rule of structures such as companies, the government, schools, army, etc. Social networks are also important in determining how diseases spread [2] ,which products we buy [3], how we vote [4] and so on. These fields are very interesting in their own right, but better understanding of them can have several practical consequences. Avoiding the spreading of fake news [5], better infection control [6], improving traffic and designing smart cities [7] are just a few examples of many. At this point it is useful to define two crude categories where networks are important in sociology and economy: In the first one, the network structure is a distribution or service network, in the second, the network connects different individuals (persons, companies, countries). In the former case the structure is usually a result of some kind of design, in the latter case however the formation of the network depends on the independent and often self-interested agents. This latter case is much more interesting for us. In case of modelling social and economic network formation, the models come mainly from two fundamental disciplines: the first is the theory of random networks, where the network formation is considered as a

random process, and the second is game theory, where strategic models of how networks are formed are developed. The static voting model is in the random network model family. Here we are not going to review the huge area of random networks, but just mention some important milestones. The field of random graphs was first introduced by the famous paper of Erdos and Rényi [8]. The Erdos-Rényi $ER_n(p)$ random graph has n vertices and each pair of vertices is independently connected with probability p. Despite the fact that $ER_n(p)$ is the simplest imaginable model of a random network, it has fascinating phase transition when p varies. Looking at many real world networks [9] such as social network, Internet, transport networks, biological networks, etc., we can see that their degree distribution is power-law, and they have the so called small-world property. Since the ER model has Poisson degree distribution, it is not a good model of real world networks. There are many extensions of this random graph model, for example the inhomogeneous version of the ER model [9] that we will use later in this paper, the configuration model [9], the generalized random graph model [9], and the most general form of the inhomogeneous random graphs [10] defined by kernel functions. In this paper we only deal with static models, but we have to mention one network growth model family: the class of preferential attachment models are such that new elements are more likely to attach to elements with high degree compared to elements with small degree. This phenomenon was first published by Barabási and Albert [11] and their model is called Barabási - Albert (BA) model. Networks generated by the BA model show power-law degree distribution. In this paper we continue and extend the work on static edge voting models [12]. The structure of this paper is as follows: after an introduction to the necessary notions and definitions, in Section 2 we turn to the Inhomogeneous Random Graphs. In this section we have two main points: the Chernoff-bounds and the fact that we can use the Gauss distribution to approximate the distribution of the degree of a given node in the network. In Section 3 we define the static edge voting model (SEV), introduce the Mixture of Gaussians (MOG) method to estimate the degree distribution of the graph (Subsection 3.1) and discuss how we generate the parameters of the SEV models to be able to execute numerical tests (Subsection 3.2). In Subsections 3.3, 3.4 and 3.5 we define and analyse 3 special cases of the SEV model: the proportional, the Poission and the biased voting model. We will demonstrate the usability of the MOG approximation with the proportional model, and in the following models we will use only this.

1.1 Notations and definitions

We denote the set $\{1,...,n\}$ with $[n]$. In our models there are only simple graphs. A G simple graph is given by a pair: $G = (V(G), E(G))$, where $V(G)$ is the vertex set of the G graph and $E(G)$ contains the edges of the graph. The vertices are marked with integer numbers, so the vertex set of the graph with n vertices is $V = [n]$. The degree of a node in the graph is definded as the number of neighbours of the node. For a given $a \in V$ we denote the degree of a by $d(a)$ or simply by d_a. We can enumerate the degrees of all the nodes in a sequence: $(d_1,...,d_n)$. This sequence is called the degree sequence of the G graph. We define the degree distribution of the G graph as the distribution of $d(U)$, where U is a randomly and uniformly chosen node. Therefore $d(U)$ is a random variable even in case of a deterministic graph.

In deterministic case we can express $\mathbb{P}(d(U)=k)$ as $\#\{a|a\in V, d_a=k\}/n$. If G is random we can handle d_a as a random variable, and we refer to the distribution of d_a as the distribution of the degree of the node a. We will use the abbreviation: $V_i = V\setminus\{i\}$. Now we define the inhomogeneous random graphs (IRG) as it is defined in [9]. Lets denote $P=\{p_{ij}\}$, $1\le i<j\le n$, the edge probabilities in the graph, and $IRG_n(P)$ for the inhomogeneous random graph for which the edges are drawn independently, and the probability that the edge $\{i,j\}$ is present in the graph equals p_{ij}. All the models in this paper have an equivalent IRG representation, and we will use either the original definition or the equivalent $IRG_n(P)$ depending on which one is more suitable.
In order to compare discrete probability distributions, we use two methods: total variation distance [9] and Jensen-Shannon divergence [13]. In general, for two probability measures μ and ν, the total variation distance is defined as:

$$d_{TV}(\mu,\nu)=\sup_A|\mu(A)-\nu(A)|$$

For discrete probability distributions with the same $\mathbb{X}$ support $(p_x)_{x\in\mathbb{X}}$ and $(q_x)_{x\in\mathbb{X}}$, these measures are given by $\mu(A)=\sum_{a\in A}p_a$ and $\nu(A)=\sum_{a\in A}q_a$. In this case it becomes:

$$d_{TV}(p,q)=\frac{1}{2}\sum_x|p_x-q_x|$$

The Jensen-Shannon divergence is the symmetrized and smoothed version of the Kullback–Leibler divergence, so first we have to define it. For discrete probability distributions $(p_x)_{x\in\mathbb{X}}$ and $(q_x)_{x\in\mathbb{X}}$ defined on the same probability space:

$$KL(p,q)=\sum_{x\in\mathbb{X}}p_x\log\left(\frac{p_x}{q_x}\right)$$

The Kullback–Leibler divergence is defined only if for all x, $q_x=0$ implies that $p_x=0$. Whenever p_x is zero the contribution of the corresponding term is interpreted as zero. The KL divergence is not symmetric. To make it symmetric define the Jensen-Shannon divergence as:

$$JS(p,q)=\frac{1}{2}KL(p,m)+\frac{1}{2}KL(q,m)$$

where the $(m_x)_{x\in\mathbb{X}}$ distribution is defined as: $m_x=\frac{1}{2}(p_x+q_x)$

2 Properties of Inhomogeneous Random graphs

It would be worthwhile to know the distribution of the d_i degree of an arbitrary i vertex in the IRG model. However, computationally it is very expensive to compute the proper distribution. Generally:

$$\mathbb{P}(d_i=k)=\sum_{S\subseteq V_i,|S|=k}\prod_{j\in S}p_{ij}\prod_{j\in V_i\setminus S}(1-p_{ij}) \tag{1}$$

In our preliminary paper [12] we have formalized the following proposition, and the proof is also given there. This is not a new theorem, but only the reformulation of the Chernoff bound [9] for the case of inhomogeneous random graphs.

Proposition 1. *If given a $IRG_n(P)$ model where $P = \{p_{ij}\}$, $i \leq i < j \leq n$, and d_i is the degree of the vertex i, then for all $D \in \mathbb{N}$:*

$\mathbb{P}(d_i \geq D) \leq e^{-tD} M_i(t)$ *for all* $t > 0$

and

$\mathbb{P}(d_i \leq D) \leq e^{-tD} M_i(t)$ *for all* $t < 0$

where

$$M_i(t) = \prod_{[n]\setminus i} (p_{ij} e^t + 1 - p_{ij})$$

Now, we prove the following corollary of Proposition 1:

Corollary 1. *If given a $IRG_n(P)$ model where $P = \{p_{ij}\}$, $i \leq i < j \leq n$, and d_i the degree of the vertex i, then for all $D \in \mathbb{N}$:*

$$\mathbb{P}(d_i \geq D) \leq (L(D))^{n-1} \text{ if } D > (n-1)\overline{p_i} \tag{2}$$

and

$$\mathbb{P}(d_i \leq D) \leq (L(D))^{n-1} \text{ if } D < (n-1)\overline{p_i} \tag{3}$$

where:

$$L(D) = \overline{p_i} \left(\frac{(1-\overline{p_i})\frac{D}{n-1}}{\overline{p_i}(1-\frac{D}{n-1})} \right)^{1-\frac{D}{n-1}} + (1-\overline{p_i}) \left(\frac{(1-\overline{p_i})\frac{D}{n-1}}{\overline{p_i}(1-\frac{D}{n-1})} \right)^{-\frac{D}{n-1}} \tag{4}$$

and

$$\overline{p_i} = \frac{1}{n-1} \sum_{j \in V_i} p_{ij}$$

Proof. Using the notations of Proposition 1:

$$M_i(t) = \left(\left(\prod_{[n]\setminus i} (p_{ij} e^t + 1 - p_{ij}) \right)^{\frac{1}{n-1}} \right)^{n-1}$$

Applying the inequality $(\prod x_i)^{\frac{1}{n}} \leq \frac{1}{n} \sum x_i$ and rearranging the terms, we have:

$$e^{-tD} M_i(t) \leq (L(D,t))^{n-1}$$

where:

$$L(D,t) = e^{-t\frac{D}{n-1}}\left(\overline{p_i}e^t + 1 - \overline{p_i}\right)$$

$$\frac{\partial L(D,t)}{\partial t} = -\frac{D}{n-1}e^{-t\frac{D}{n-1}}\left(\overline{p_i}e^t + 1 - \overline{p_i}\right) + e^{-t\frac{D}{n-1}}\overline{p_i}e^t =$$
$$= e^{t(1-\frac{D}{n-1})}\overline{p_i}\left(1 - \frac{D}{n-1}\right) + (\overline{p_i} - 1)\frac{D}{n-1}e^{-t\frac{D}{n-1}}$$

And the second derivative:

$$\frac{\partial^2 L(D,t)}{\partial^2 t} = e^{t(1-\frac{D}{n-1})}\overline{p_i}\left(1 - \frac{D}{n-1}\right)^2 - (\overline{p_i} - 1)\left(\frac{D}{n-1}\right)^2 e^{-t\frac{D}{n-1}}$$

Since $\frac{\partial^2 L(D,t)}{\partial^2 t}$ is positive for all t therefore $L(D,t)$ is convex for any fixed D, so $L(D,t)$ has a unique minimum, what we can get solving $\frac{\partial L(D,t)}{\partial t} = 0$ for t:

$$t^* = argmin_t L(D,t) = ln\left(\frac{(1-\overline{p_i})\frac{D}{n-1}}{\overline{p_i}(1-\frac{D}{n-1})}\right)$$

After substitution back to $L(D,t)$:

$$L(D) = L(D,t^*) = \overline{p_i}\left(\frac{(1-\overline{p_i})\frac{D}{n-1}}{\overline{p_i}(1-\frac{D}{n-1})}\right)^{1-\frac{D}{n-1}} + (1-\overline{p_i})\left(\frac{(1-\overline{p_i})\frac{D}{n-1}}{\overline{p_i}(1-\frac{D}{n-1})}\right)^{-\frac{D}{n-1}}$$

We still have to show that in case of (2) $t^* > 0$ and similarly in case of (3) $t^* < 0$. Lets start with (2) and assume that $\frac{D}{n-1} > \overline{p_i}$

$$t^* = ln\left(\frac{(1-\overline{p_i})\frac{D}{n-1}}{\overline{p_i}(1-\frac{D}{n-1})}\right) > 0$$

if and only if

$$(1-\overline{p_i})\frac{D}{n-1} > \overline{p_i}\left(1 - \frac{D}{n-1}\right)$$

After rearrange the terms, we get:

$$\frac{D}{n-1} > \overline{p_i}$$

Similarly, if we assume that $\frac{D}{n-1} < \overline{p_i}$ we can see that $t^* < 0$ in the same way.

□

Corollary 1 is a less strict, but easier to use theorem when $L(D) < 1$ holds. In ref [12], we showed that we can find an approximation of the degree distribution of node i with the $\mathcal{N}(\mu,\sigma)$ Gauss distribution, setting the parameters to $\mu = \mathbb{E}[d_i] = \sum_{j\in V_i} p_{ij}$ and $\sigma^2 = Var[d_i] = \sum_{j\in V_i} p_{ij}(1-p_{ij})$. The Lyapunov variant of the central limit theorem [14] gives the theoretical background for this. In this paper we refer to this approximation method as the Gauss approximation of the degree distribution of the nodes.

3 The static edge voting model

We divide the set of agents to disjoint groups: $V = [N] = S_0 \cup ... \cup S_M$, and suppose that the behaviour of the agents belonging to a group is the same. Denote V_{ab} a random variable that means the vote of the a node to the $\{a,b\}$ edge. For all $a \in S_i$ and $b \in S_b$, V_{ab} has the same probability distribution. For any different nodes a and b the probability that there will be an edge between the nodes a and b in the graph is up to the incoming votes and it is equal to $s(V_{ab}, V_{ba})$, where s is the edge probability function.

Definition 1. *Denote $SEV_N(\ \{S_0,...,S_M\},\ \{F_{ij}|i,j \in \{0,...,M\}\},\ s)$ the static edge voting model, where:*

- *N is the number of agents and $V = [N]$ is the set of agents.*
- *$\{S_0,...,S_M\}$ is a partition of the V set and M is the number of groups in the partition. (Here we allow empty sets in the partition.)*
- *Denote V_{ab} the random variable that means the vote of the a node to the $\{a,b\}$ edge candidate.*
- *$\{F_{ij}|i,j \in \{0,...,M\}\}$ is a set of probability distributions. For all $a \in S_i$ and $b \in S_b$ the law of V_{ab} is F_{ij}.*
- *$s : \mathbb{R} \times \mathbb{R} \mapsto [0,1]$ is the edge probability function, a non-decreasing integrable function in both arguments. Every $\{a,b\}$ $(a \neq b)$ node pair will be part of the graph with probability $s(V_{ab}, V_{ba})$.*

We proved in [12] the following simple proposition:

Proposition 2. *Given a $SEV_N(\ \{S_0,...,S_M\},\ \{F_{ij}|i,j \in \{0,...,M\}, s\}$ static edge voting model the expected value of the d_a degree of the a node, if $a \in S_i$:*

$$\mathbb{E}[d_a | a \in S_i] = \sum_{j=0}^{M} k_j \mathbb{E}[s(V_{ab}, V_{ba}) | b \in S_j] - \mathbb{E}[s(V_{ab}, V_{ba}) | b \in S_i, a \neq b]$$

where $k_i = |S_i|$

3.1 Approximation of the degree distribution

It is clear that the equivalent $IRG_n(P)$ of the $SEV_N(\ \{S_0,...,S_M\},\ \{F_{ij}|i,j \in \{0,...,M\}\},\ s)$ is the *IRG* with edge probabilities:

$P_{ab} = \mathbb{P}(\{a,b\} \in E)$

We can use the total probability rule to express the degree distribution of the SEV_N random network:

$$\mathbb{P}(d(U) = k) = \sum_{i=0}^{M} \mathbb{P}(U \in S_i) \mathbb{P}(d(U) = k | U \in S_i) \tag{5}$$

We can get $\mathbb{P}(U \in S_i) = \frac{\#S_i}{N}$ directly from the SEV_N model, and we can compute $\mathbb{P}(d(U) = k|U \in S_i)$ directly using the equivalent $IRG_n(P)$ model and equation (2). However, the direct use of equation (2) is computationally very expensive. In spite of this, we can use the Mixture of Gaussian's (MOG) model: apply the Gauss approximation of $\mathbb{P}(d(U) = k|U \in S_i)$ as defined in Section 2. For arbitrary $a \in V$ denote $\mu_a = \sum_{b=1}^{N} P_{ab}$ and $\sigma_a^2 = \sum_{b=1}^{N} (P_{ab}(1-P_{ab}))$. It is easy to see that for a fixed S_i and for arbitrary $a, b \in S_i$, $P_{ac} = P_{bc}$ for all $c \in V$. Therefore the μ_a and σ_a^2 values are the same for all $a \in S_i$, that we denote by μ_{S_i} $\sigma_{S_i}^2$. We use the estimation $\mathbb{P}(d(U) = k|U \in S_i) \approx F(k+0.5; \mu_{S_i}, \sigma_{S_i}^2) - F(k-0.5; \mu_{S_i}, \sigma_{S_i}^2)$, where $F(x; \mu_{S_i}, \sigma_{S_i}^2)$ is the cumulative distribution function of the Gauss distribution with parameters μ_{S_i} and $\sigma_{S_i}^2$. Plugging these back to equation (5), we have

$$\mathbb{P}(d(U) = k) \approx \sum_{i=0}^{N-1} \mathbb{P}(U \in S_i) \left(F(k+0.5; \mu_{S_i}, \sigma_{S_i}^2) - F(k-0.5; \mu_{S_i}, \sigma_{S_i}^2) \right) \tag{6}$$

3.2 Degree sequences

In the rest of the paper we would like to present numerical examples, and for this, we need to generate $D = (D_1, ..., D_n)$ degree sequences. We use four methods to generate these sequences: constant, uniform, binomial and lognormal. The constant degree sequence is the simplest: we just create a N-length sequence of constant values: $ConstSeq(C, N) = (C, ..., C)$. The second method is the uniform. We denote this with $UniformSeq(MIN, MAX, N)$, where MIN, MAX and N are the minimum, the maximum and the size of the sequence. The last two are the binomial and lognormal, denoted by $BinomSeq(p, N)$ and $LognormalSeq(\mu, \sigma^2, N)$, where p and N are the parameters of the binomial distribution and N is the length of the sequence at the same time, while μ and σ^2 are the parameters of the lognormal distribution. We omit here the algorithm what we used generating these sequences. The main point here is that the values in the generated sequences are approximately obey the uniform, lognormal and binomial laws. After the degree sequence generation, we computed the empirical distribution of each sequence. In case of the sequences $BinomSeq(0.1, 25)$, $LognormalSeq(1.7, 0.4, 25)$, $BinomSeq(0.1, 1000)$, $LognormalSeq(4, 0.6, 1000)$ the result distributions are plotted on Figure 1. Note that except the constant sequence case, the empirical distributions of the sequences are just the approximations of the source distributions.

3.3 The proportional edge voting model

The static proportional edge voting model is a static random network model where the V_{ab} random variables do not depend on b, and for all $a \in S_i$, $V_{ab} \sim Bernoulli(\frac{i}{N-1})$ and $s(x) = x/2$. This model can be interpreted as that the agent a who belongs to the group S_i would like to achieve $D_a = i$ degree. The probability that the a agent gives a vote on the $\{a, b\}$ edge candidate is $\frac{D_a}{N-1}$. This voting rule justifies the "proportional" name. The direct consequence of this setup is that the number of agent groups $M = N - 1$, and S_i is the group of agents who target degree is i. The general definition of the static edge voting model assumes that the distribution of the

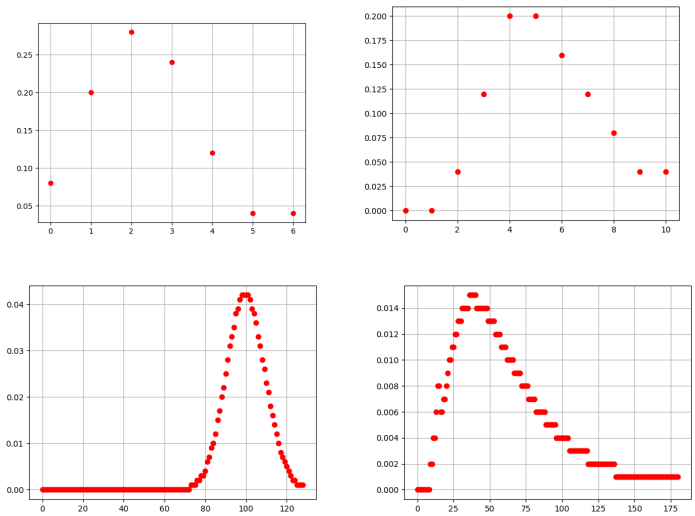

Figure 1
The empirical distribution of the degree sequences: $BinomSeq(0.1,25)$ (top-left), $LognormalSeq(1.7,0.4,25)$ (top-right), $BinomSeq(0.1,1000)$ (bottom-left),$LognormalSeq(4,0.6,1000)$ (bottom-right)

V_{ab} random variables are given. In case of the proportional model this is equivalent with that the $\{D_a|a \in V\}$ degree sequence is given, where D_a can be considered as the desired degree of agent a. It is clear that:

$$\mathbb{P}(\{a,b\} \in E) = \frac{1}{2}\frac{D_a + D_b}{N-1}$$

From Proposition 2 we can directly get:

$$\mathbb{E}[d_a|a \in S_i] = \frac{1}{2(N-1)}\left(i(N-2) + \sum_{j=0}^{N-1} k_j j\right)$$

Using the estimation: $N-1 \approx N$ and $N-2 \approx N$ we have:

$$\mathbb{E}[d_a|a \in S_i] \approx \frac{1}{2N}\left(iN + \sum_{j=0}^{N-1} k_j j\right) = \frac{1}{2}\left(i + \sum_{j=0}^{N-1} \frac{k_j}{N} j\right) = \frac{1}{2}\left(i + \overline{D}\right) \tag{7}$$

Where $\overline{D} = \sum_{j=0}^{N-1} \frac{k_j}{N} j$ is the mean target degree. This result shows that this model has a "smoothing" effect on the degrees. Equation 7 also shows that if an agent a has a desired degree D_a close to the average $\overline{D}$ desired degree, then she/he can easy achieve D_a in mean, however when D_a is far from $\overline{D}$, then hitting the aims is much harder: the $\overline{D}$ average value has a gravity.
Lets continue with the degree distribution of the static edge voting model. For

this, use the four degree sequence generator methods to generate the desired degree sequences. To be able to compute the exact degree distributions using equations (2) and (5) first we use a small network with 25 nodes. The desired degree sequences are: $ConstSeq(2,25)$, $BinomSeq(0.1,25)$, $LognormalSeq(1.7,0.4,25)$ and $UniformSeq(0,4,25)$. The degree distribution of each case is plotted on Figure 2. On Figure 3 we also plotted the MOG approximation of the degree distributions. In order to compare the exact degree distributions and the MOG approximations, we computed the total variation distance and the Jensen-Shannon divergence of the distributions. The results are summarised in Table 1. We can conclude that the exact and the MOG approximated distributions are close to each other. We can not compute the exact degree distribution for greater networks, because it is computationally very expensive. However, in one special case, we know the exact distribution: in case of $ConstSeq(C,N)$ desired degree sequence the probability of each edge is $p(N,C) = \frac{C}{N-1}$, therefore this model is equivalent to the Erdos-Renyi random graph: $ER(N,p(N,C))$. We know that the degree distribution of this ER model is binomial with parameters $N-1$ and $p(N,C)$. We computed therefore the total variation distance and the Jensen-Shannon divergence of the binomial distribution with parameters $N-1$ and $p(N,C)$ and the MOG approximation of the degree distribution with desired degree sequence $ConstSeq(C,N)$, where $N \in \{10,20,...,2000\}$ and $C = N/10$. The results are in Figure 4. We can see, that both distances decrease as N increases, however the rate of decline is decreasing. We can consider the proportional edge voting model as a transformation of the distribution of the input parameters, e.g. the distribution of the desired degree sequence. To be able to illustrate this with numerical examples, we set the desired degree sequence to $ConstSeq(100,1000)$, $LognormalSeq(4,0.6,1000)$, $BinomSeq(0.1,1000)$ and $UniformSeq(0,99,1000)$. We used the MOG approximation of the degree distribution, and the results are in Figure 5. We also computed the empirical mean and variance of the desired degree distribution and the approximated degree distribution of the proportional model. These values are collected in Table 2. From Equation (7) we may expect that the model has a shrinking effect to the desired degree distribution. However, this phenomenon is valid only in case of the lognormal and uniform distributions. The constant degree sequence is a special case, but it is not clear why this effect does not work when the input degree distribution is binomial. Finally, we apply Corollary 1 to the IRG equivalent with the proportional edge voting model to bound the probability that the degree of a given node is greater/smaller than a value. In this numerical example, we set the desired degree sequence to $LognormalSeq(4,0.6,1000)$ and apply Corollary 1 to the nodes with degrees 15, 60, 130, 170 respectively. Note that the average desired degree in this case is 61.2919. On Figure 6 we plotted the bound probability in the function of D. In the case, when the desired degree of the node is 15, we can realize that the bound probabilities have a maximum around 38. This observation meets with equation (7). We can do the same observation when the desired degree of the the node is 60, 130 or 170. Looking at the shape of the curves, we may have the idea, that the Chernoff bound and the Gauss approximation of the degree distribution of a given node give similar results.

Table 1
Total variation distance (TV) and Jensen-Shannon divergence (JS) of the exact degree distributions and the MOG approximation of the degree distributions when the desired degree sequences are: $ConstSeq(2,25)$, $BinomSeq(0.1,25)$, $LognormalSeq(1.7,0.4,25)$ and $UniformSeq(0,4,25)$

	ConstSeq	BinomSeq	LognormalSeq	UniformSeq
TV	0.0668770019	0.0459734622	0.0257139778	0.0535480364
JS	0.0035851277	0.0015726680	0.0008468790	0.0022105596

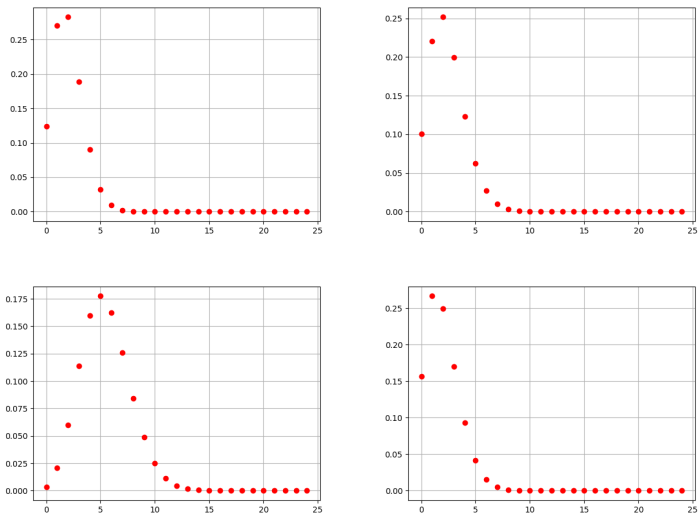

Figure 2
The exact degree distributions of the proportional edge voting model when the desired degree sequences are: $ConstSeq(2,25)$ (top-left), $BinomSeq(0.1,25)$ (top-right), $LognormalSeq(1.7,0.4,25)$(bottom-left), $UniformSeq(0,9,25)$ (bottom-right)

Table 2
The mean and the variance of the desired degree distribution (μ_D, σ_D) and the degree distribution of the proportional edge voting model using the Mixture of Gaussians approximation ($\mu_{d(U)}$, $\sigma_{d(U)}$)

	ConstSeq	BinomSeq	LognormalSeq	UniformSeq
μ_D	100	99.88	61.2919	49.5
σ_D	0	87.8295	1242.1627	833.25
$\mu_{d(U)}$	100	99.88	61.292	49.5
$\sigma_{d(U)}$	90.0733	111.8468	366.913	254.6096

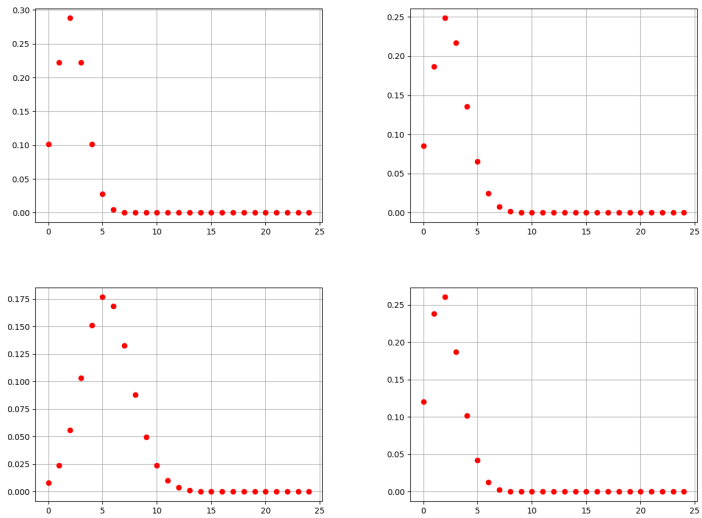

Figure 3
The degree distributions of the proportional edge voting model computing with the MOG approximation, when the desired degree sequences are: $ConstSeq(2,25)$ (top-left), $BinomSeq(0.1,25)$ (top-right), $LognormalSeq(1.7,0.4,25)$(bottom-left), $Uniform(0,9,25)$ (bottom-right)

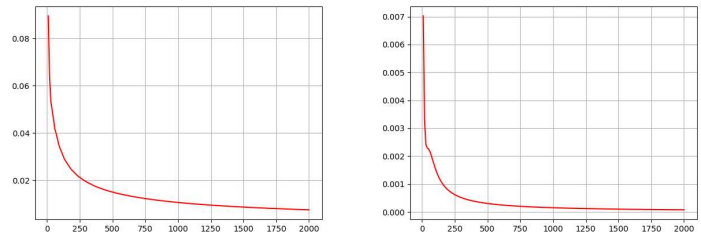

Figure 4
Total variation distance (left) and the Jensen-Shannon divergence (right) of the binomial distribution with parameters $N-1$ and $p(N,C)$ and the MOG approximation of the degree distribution with desired degree sequence $ConstSeq(C,N)$, where $N \in \{10,20,...,2000\}$ and $C = N/10$

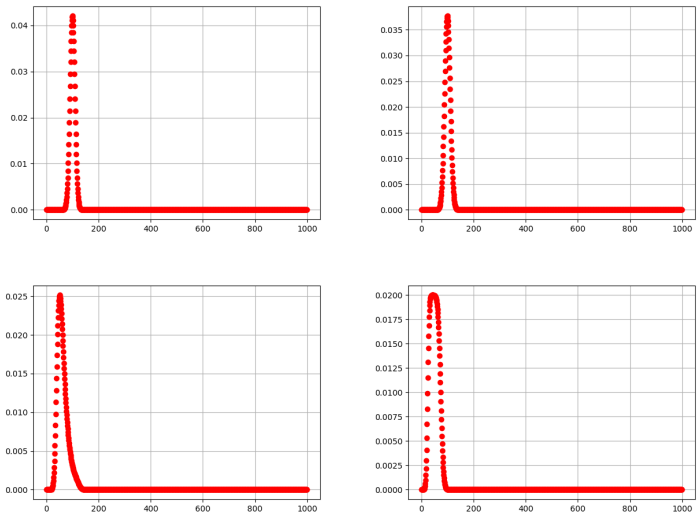

Figure 5
The degree distributions of the proportional edge voting model computing with the Mixture of Gaussians approximation, when the desired degree sequences are: $ConstSeq(100, 1000)$ (top-left), $BinomSeq(0.1, 1000)$ (top-right), $LognormalSeq(4, 0.6, 1000$(bottom-left), $Uniform(0, 99, 1000)$ (bottom-right)

.

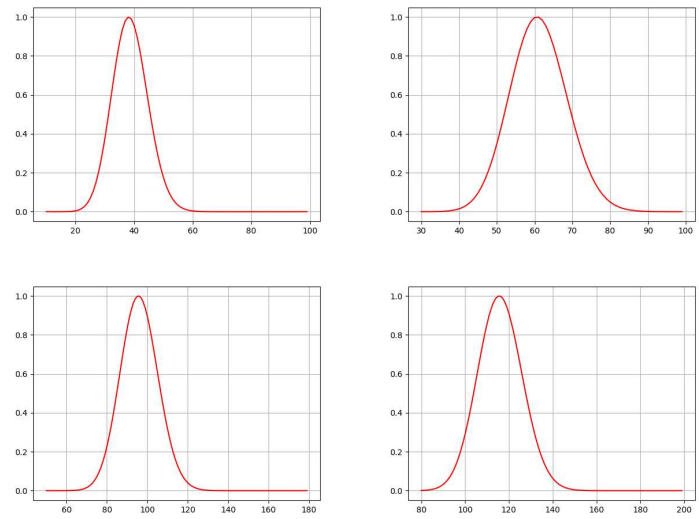

Figure 6
Applying the additive version of the Chernoff bound (Corollary 1) to the IRG equivalent with the proportional edge voting model with $LognormalSeq(4, 0.6, 1000)$ desired degree sequence. We plotted the bound probability in the function of D. The desired degree of the selected nodes are 15 (top-left), 60 (top-right),130 (bottom-left) and 170 (bottom-right).

3.4 Voting model with Poisson voting law

Let's discuss an example when the nodes may have more then one vote to a link. Consider the case when $V_{ab} \sim$ Poisson (λ_i) if $a \neq b$ and $a \in S_i$ and the edge probability function is given as $s(x) = 1 - e^{-\lambda x}$, where λ is the control parameter. We have shown [12] that the probability of the $\{a,b\}$ edge is

$$\mathbb{P}(\{a,b\} \in E | a \in S_i, b \in S_j) = 1 - e^{(\lambda_i+\lambda_j)(e^{-\lambda}-1)} \tag{8}$$

And the expected degree of the a node is:

$$\mathbb{E}[d_a | a \in S_i] = N - 1 + e^{2\lambda_i(e^{-\lambda}-1)} - \sum_{j=0}^{N-1} e^{(\lambda_i+\lambda_j)(e^{-\lambda}-1)} \tag{9}$$

The parameters of the Poission voting model are $\{\lambda_i | i \in V\}$ and the λ control parameter. We would like to compare the behaviour of the Poission model to the proportional model. To be able to do this we use the same sequences to set the λ_i parameters of the model: $ConstSeq(100, 1000)$, $BinomSeq(0.1, 1000)$, $LognormalSeq(4$,0.6, 1000) and $UniformSeq(0, 99, 1000)$, but in this case we apply these values to set the λ_i parameters. However, we have an additional model parameter: λ, that we have to set somehow. It is reasonable to set λ such that the mean of the $\{\lambda_i\}$ parameters approximately equals to the mean of the (MOG approximated) degree distribution of the Poisson model. To achieve this, we used the bisection [15] root-finder to find the appropriate parameter. The fitted parameters are collected in Table 3. We also added the variance of the degree distribution of the Poisson model and the proportional edge voting model with the same $\{D_i\}$ sequence. The degree distribution of the Poisson models with the fitted parameter are plotted on Figure 7. According to the plots and the variance numbers we can conclude that the Poisson voting model and the proportional voting models behave in a very similar way in this setup.
We can do the same, but fitting to the variance. The results of this approach are in Table 4. We can see that changing the parameter λ we can get similar variability as the input $\{\lambda_i\}$ parameters have, but the price is that the mean is shifted.
We can now ask what is the role of the λ parameter in the model. We can investigate the effect if we fix the $\{\lambda_i\}$ parameters and observe how the mean and the variance is changing when we change λ. We expect that λ behaves as the accelerator of the model, as it is increasing, the mean is increasing, however the variance will reach a maximum and then it will fall. Our experiments confirm these expectations. Fixing the $\{\lambda_i\}$ parameters in $LognormalSeq(4, 0.6, 1000)$, we plotted the results in Figure 8.

3.5 Biased voting model

In this variant of the static edge voting model we modify the proportional edge voting model. Here we consider the D_a values as social capital, where the greater D_a means not only more desired degree but also greater influence. This idea is similar to the preferential attachment rule in the Barabasi-Albert model [11]. To be

Table 3
The fitted λ parameter of the Poisson voting model, when the $\{\lambda_i\}$ parameters setted to $ConstSeq(100, 1000)$, $BinomSeq(0.1, 1000)$, $LognormalSeq(4, 0.6, 1000)$ and $UniformSeq(0, 99, 1000)$. The first and the second lines are the mean and the variance of the λ_i parameters. The third line is the fitted λ parameter for each cases. The forth and the fifth lines are the variance of the (approximated) degree distribution in case of the Poisson and the proportional voting model respectively.

	ConstSeq	BinomSeq	LognormalSeq	UniformSeq
μ_{λ_i}	100	99.88	61.2919	49.5
σ_{λ_i}	0	87.8295	1242.1627	833.25
λ	0.0005280	0.00052802	0.00051923	0.00051572
$\sigma^{Poi}_{d(U)}$	90.1494	109.7385	345.0582	246.0220
$\sigma^{pro}_{d}(U)$	90.0733	111.8468	366.913	254.6096

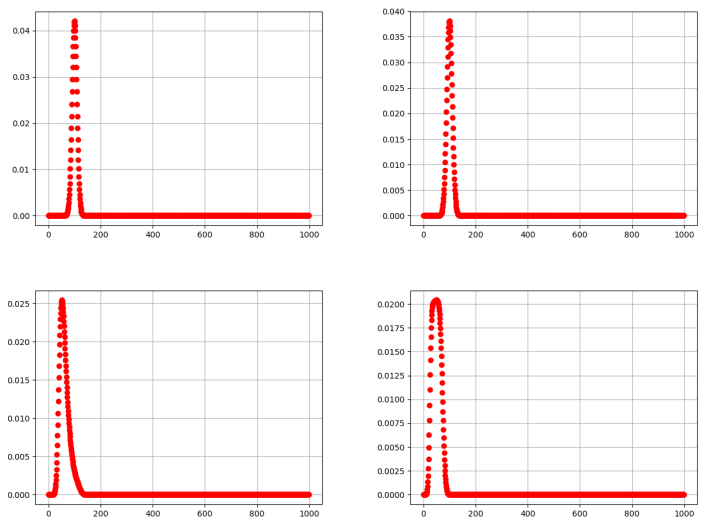

Figure 7
The degree distribution of the Poisson voting model, when the $\{\lambda_i\}$ parameters are $ConstSeq(100, 1000)$, $BinomSeq(0.1, 1000)$, $LognormalSeq(4, 0.6, 1000)$ and $UniformSeq(0, 99, 1000)$ and the λ parameter of each cases is setted to 0.0005280, 0.00052802, 0.00051923, 0.00051572 respectively

Table 4
The fitted λ parameter of the Poisson voting model, when the λ_i parameters setted to $ConstSeq(100,1000)$, $BinomSeq(0.1,1000)$, $LognormalSeq(4,0.6,1000$ and $UniformSeq(0,99,1000)$. The first line is the variance of the λ_i parameters. The second line is the fitted λ parameter for each cases. The third and the forth lines are the means of the (approximated) degree distribution in case of the Poisson and the proportional voting model respectively.

	ConstSeq	BinomSeq	LognormalSeq	UniformSeq
σ_{λ_i}	0	87.8295	1242.1627	833.25
λ	0.05850230	0.00042395	0.00112025	0.00105202
$\mu_{d(U)}^{Poi}$	998.999	81.089	126.781	97.933
$\mu_{d(U)}^{pro}$	100	99.88	61.2919	49.5

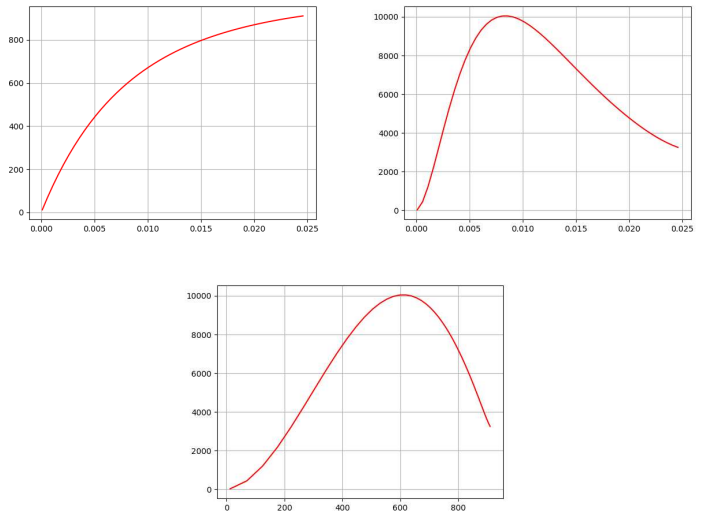

Figure 8
Above, the mean and the variance of the Poisson voting model are plotted in the function of parameter λ, when the λ_i parameters are $LognormalSeq(4,0.6,1000)$. Below the variance is plotted in the function of the mean

able to formalize this with the edge probability function, first define:

$$v(V_{ab}, V_{ba}) = \frac{D_a}{N-1} V_{ab} + \frac{D_b}{N-1} V_{ba}$$

In the function v the vote of agent a has a weight $\frac{D_a}{N-1}$. Using the v function we can formalize the s function for any fixed $\eta > 0$:

$$s_b(V_{ab}, V_{ba}) = 1 - e^{-\eta v(V_{ab}, V_{ba})}$$

As voting rule we use the Bernoulli voting rule, where V_{ab} does not depend on b and $\mathbb{P}(V_{ab} = 1) = p_i$ if $a \in S_i$

Proposition 3. *In the static edge voting model using the Bernoulli voting rule and the s_b edge probability function, for any $a \in V$ node:*

$$\mathbb{P}[\{a,b\} \in E] = p^*_{ab} + p^*_{ba} - p^*_{ab} p^*_{ba} \tag{10}$$

where:

$$p_{ab} = \mathbb{P}(V_{ab} = 1)$$

$$p^*_{ab} = p_{ab} \left(1 - e^{-\eta \frac{D_a}{N-1}}\right)$$

Proof. Using the definition of s_b and the fact that the expected value of a Bernoulli random variable is its parameter, we have:

$$\mathbb{P}[\{a,b\} \in E] = \mathbb{E}[1 - e^{-\eta v(V_{ab}, V_{ba})}] = 1 - \mathbb{E}[e^{-\eta v(V_{ab}, V_{ba})}]$$

Since the agents can give 0 or 1 vote, we can express the expected value directly using the definition:

$$1 - \mathbb{E}[e^{-\eta v(V_{ab}, V_{ba})}] = 1 - e^0 (1 - p_{ab})(1 - p_{ba}) - e^{-\eta \frac{D_a}{N-1}} p_{ab}(1 - p_{ba}) -$$
$$- e^{-\eta \frac{D_{ba}}{N-1}} (1 - p_{ab}) p_{ba} - e^{-\eta \frac{D_a + D_b}{N-1}} p_{ab} p_{ba} =$$

After doing the multiplications, rearranging and grouping the appropriate terms, we get the statement:

$$= p_{ab}(1 - e^{-\eta \frac{D_a}{N-1}}) + p_{ba}(1 - e^{-\eta \frac{D_b}{N-1}}) - p_{ab}(1 - e^{-\eta \frac{D_a}{N-1}}) p_{ba}(1 - e^{-\eta \frac{D_b}{N-1}}) =$$
$$= p^*_{ab} + p^*_{ba} - p^*_{ab} p^*_{ba}$$

□

Proposition 4. *In the static edge voting model using the Bernoulli voting rule and the s_b edge probability function, for any $a \in V$ node:*

$$\mathbb{E}[d_a] = \sum_{b \in V_a} (p^*_{ab} + p^*_{ba} - p^*_{ab} p^*_{ba})$$

where:

$$p_{ab} = \mathbb{P}(V_{ab} = 1)$$

$$p^*_{ab} = p_{ab} \left(1 - e^{-\eta \frac{D_a}{N-1}}\right)$$

Proof. Since the degree of a node is the number of its neighbours in the graph, and the expected value of a Bernoulli random variable is its parameter:

$$\mathbb{E}[d_a] = \sum_{b \in V_a} \mathbb{P}[\{a,b\} \in E]$$

Applying Proposition 3 we get the statement. □

Using equation (10) we can construct the equivalent IRG model. As before in the case of the Poisson model, we would like to compare the model to this proportional model and to the Poisson model. In order to do this we use the same sequences to set the $\{D_i\}$ parameters of the model: $ConstSeq(100, 1000)$, $BinomSeq(0.1, 1000)$, $LognormalSeq(4, 0.6, 1000)$ and $UniformSeq(0, 99, 1000)$. We use the proportional voting rule: $p_i = \frac{D_i}{N-1}$ but we also have to set the η parameter. We do the same as before: set the η parameter such that the mean of the $\{D_i\}$ parameters approximately equals to the mean of the (MOG approximated) degree distribution. To achieve this, we used the bisection [15] root-finder again. The fitted parameters are collected in Table 5. and the corresponding degree distributions are on Figure 9. As in the case of the Poission model, we also fitted the η parameters to hit the variance of the $\{D_i\}$ parameter set. The results of this method are in Table 6. When we fitted the η parameter to hit the mean of the input $\{D_i\}$ sequence, we can see that we can achieve greater variance in the degree distribution than in the case of the other two models. While we fitted to hit the variance, we can now see that the shift of the mean is not so high then in case of the Poisson model.
Now we investigate the role of the η parameter with numerical experiments. As before, fix the $\{D_i\}$ parameters to $LognormalSeq(4, 0.6, 1000)$ and compute the mean and the variance of the approximated degree distribution for each $\eta \in 0, 1, ..., 100$ The results are plotted in Figure 10.

4 Discussion

In this paper we continued and extended our preliminary work [12] about static edge voting models. In Section 2 started with reviewing the properties of inhomogeneous random graphs (IRG), because IRGs are our basic tool, and proved a corollary of the Chernoff-bounds what we can use to bound the probability that the degree of a fixed node is below or above of a given D value. In Section 3 we gave

Table 5
The fitted η parameter of biased model, when the D_i parameters setted to $ConstSeq(100, 1000)$, $BinomSeq(0.1, 1000)$, $LognormalSeq(4, 0.6, 1000$ and $UniformSeq(0, 99, 1000)$. The first and the second lines are the mean and the variance of the D_i parameters. The third line is the fitted η parameter for each case. The next lines are the variance of the (approximated) degree distribution in case of the biased, the proportional voting model and the Poisson model respectively.

	ConstSeq	BinomSeq	LognormalSeq	Uniform
μ_{D_i}	100	99.88	61.2919	49.5
σ_{D_i}	0	87.8295	1242.1627	833.25
η	7.1914	7.1591	9.5026	11.1728
$\sigma^b_{d(U)}$	90.0733	148.266	884.0557	444.9441
$\sigma^{pro}_d(U)$	90.0733	111.8468	366.913	254.6096
$\sigma^{Poi}_{d(U)}$	90.1494	109.7385	345.0582	246.0220

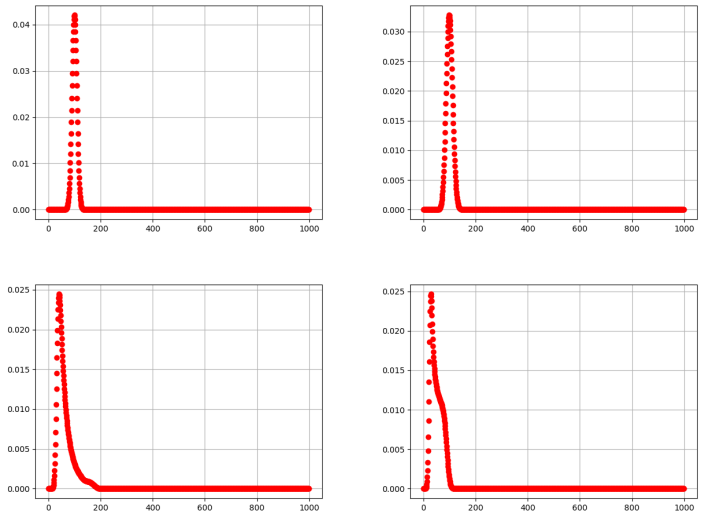

Figure 9
The degree distribution of biased voting model, when the D_i parameters are $ConstSeq(100, 1000)$, $BinomSeq(0.1, 1000)$, $LognormalSeq(4, 0.6, 1000$ and $UniformSeq(0, 99, 1000)$ and the η parameter of each cases is setted to 7.1914, 7.1591, 9.5026, 11.1728 respectively

Table 6
The fitted η parameter of the biased voting model, when the D_i parameters setted to $ConstSeq(100, 1000)$, $BinomSeq(0.1, 1000)$, $LognormalSeq(4, 0.6, 1000$ and $UniformSeq(0, 99, 1000)$. The first line is the variance of the λ_i parameters. The second line is the fitted η parameter for each case. The next lines are the means of the (approximated) degree distribution in case of the biased, the proportional voting model and the Poission model respectively.

	ConstSeq	BinomSeq	LognormalSeq	UniformSeq
σ_{D_i}	0	87.8295	1242.1627	833.25
η	0	3.8873	16.348	26.3628
$\mu^{b}_{d(U)}$	0	63.654	82.00028	76.305
$\mu^{pro}_{d(U)}$	100	99.88	61.2919	49.5
$\mu^{Poi}_{d(U)}$	998.999	81.089	126.781	97.933

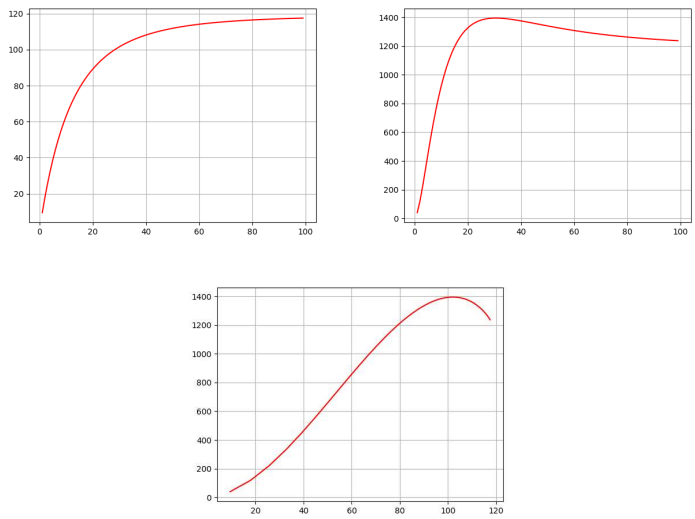

Figure 10
Above, the mean and the variance of the biased voting model are plotted in the function of parameter η, when the D_i parameters are $LognormalSeq(4, 0.6, 1000)$ Below the variance is plotted in the function of the mean

the general definition of the static edge voting models, and introduced the method of Mixture of Gaussians to be able to estimate the degree distribution of the model. We dealt with three special cases: the proportional edge voting model, the Poisson model and the biased voting model. In case of the proportional model any a agent votes for the $\{a,b\}$ edge candidate with probability proportional to its D_a desired degree, and the probability that $\{a,b\}$ becomes an edge is just the average of the incoming votes. We have seen, that the proportional model has a smoothing effect, the expected degree of a node a is approximately the average of the D_a desired degree of a and the $\overline{D}$ mean desired degree. In the Poisson model the agents can give more votes to an edge candidate, and the distribution of the votes given by agent a is Poisson(λ_i) if $a \in S_i$. The edge probability function here is $s(x) = 1 - e^{-\lambda x}$ where λ is a control parameter. The input parameters are in this model the $\{\lambda_a | a \in V\}$ parameters of the Poisson distributions and the λ control parameter. We have seen, that if we set λ such that the mean of the degrees is equal to the mean of $\{\lambda_a | a \in V\}$, then the Poisson edge voting model behaves like the proportional model, however we can get similar variance in the degrees like in the $\{\lambda_a | a \in V\}$ input sequence at the price that the mean of the degree distribution is shifting. The biased voting model is the modification of the proportional model in a way that the votes of agent a is weighted by $\frac{D_a}{N-1}$. In this model D_a can be considered as social capital. This biased behaviour is enforced by the edge probability function: $s_b(V_{ab}, V_{ba}) = 1 - \exp\left(-\eta\left(\frac{D_a}{N-1}V_{ab} + \frac{D_b}{N-1}V_{ba}\right)\right)$. We have seen that we can achieve greater variability in degree as in case of the proportional and the Poission models. Lets do now a final comparison as a summary. Fix again the input parameters of all the three models: in case of the proportional and biased models set the $\{D_i | i \in V\}$ sequences to $LognormalSeq(4, 0.6, 1000)$, and in case of the Poisson model set $\{\lambda_i | i \in V\}$ to $LognormalSeq(4, 0.6, 1000)$ again. Choose the control parameters λ and η in case of the Poisson and the biased models forcing the mean of the degree distribution equal to the mean of the input $LognormalSeq(4, 0.6, 1000)$ sequence. Now compute the expected degree of the nodes for a reference node for each non-empty S_i agent group, and plot $\mathbb{E}[d_a | a \in S_i]$ in the function of the parameter (D_i or λ_i). The results are in Figure 11. We can see that the proportional and the Poission models are behaving almost identically, and we can experience the "smoothing" behaviour in both cases. However, in case of the biased model after a short initial period the curve of the expected degrees becomes close to the identical function, which shows that the biased model more-or-less preserves the variability of the input parameter set in the degree distribution.

We note that the SEV model better describes the process in which agents explore their environment in the social space, but the survival of relationships are affected by other factors. A natural extension is to make the model temporal, and the previous note makes sense only in temporal networks: agents first explore their environment by making connections, but maintaining these connections is a different process. Finally, here we supposed a small population, where every agent can link to any other agent. However, in larger populations this is not true, so we have to include the distance into our framework in some way.

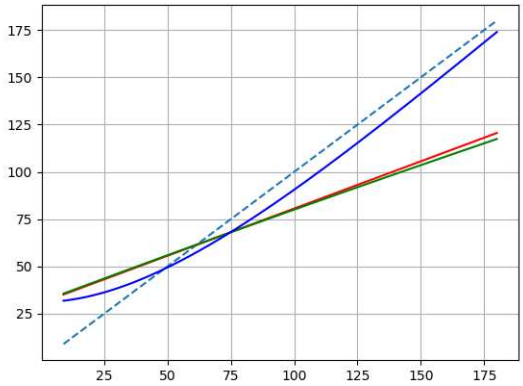

Figure 11
Comparison of the expected degrees of the nodes in the function of the model parameters. The horizontal axis is the parameter and the vertical axes is the expected degree of the node in the function of the parameter. The proportional, the Poission and the biased models are represented by the red, green and blue lines respectively. The blue dashed line is the identical function.

References

[1] Anna Zafeiris and Tamás Vicsek. *Why We Live in Hierarchies?: A Quantitative Treatise*. Springer, 2017.

[2] Albert-László Barabási et al. *Network science*. Cambridge university press, 2016.

[3] Everett M Rogers. Diffusion of innovations. free press. *New York*, 551, 2003.

[4] Margarita Vázquez-Brage, Ignacio Garcıa-Jurado, and Francesc Carreras. The owen value applied to games with graph-restricted communication. *Games and Economic Behavior*, 12(1):42–53, 1996.

[5] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. *Science*, 359(6380):1146–1151, 2018.

[6] A Barrat, M Barthélemy, and A Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2012.

[7] Tamás Péter and Krisztián Szabó. Combined mathematical modeling of different transport networks, considerations and complex analysis. *Acta Polytechnica Hungarica*, 14(2):7–26, 2017.

[8] Paul Erdos and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*, 5:17–61, 1960.

[9] Remco Van Der Hofstad. *Random graphs and complex networks*, volume 1. Cambridge University Press, 2016.

[10] Béla Bollobás, Svante Janson, and Oliver Riordan. The phase transition in inhomogeneous random graphs. *Random Structures & Algorithms*, 31(1):3–122, 2007.

[11] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.

[12] Róbert Pethes and Levente Kovács. Static edge voting models. In *IEEE 18th International Symposium on Computational Intelligence and Informatics (CINTI 2018) Budapest, Hungary*, pages 223–228. IEEE, 2018.

[13] Bent Fuglede and Flemming Topsoe. Jensen-shannon divergence and hilbert space embedding. In *International Symposium onInformation Theory, 2004. ISIT 2004. Proceedings.*, page 31. IEEE, 2004.

[14] Hans Fischer. *A history of the central limit theorem: From classical to modern probability theory*. Springer Science & Business Media, 2010.

[15] D Vaughan Griffiths and Ian Moffat Smith. *Numerical methods for engineers*. Chapman and Hall/CRC, 2006.

Classification of Special Web Reviewers Based on Various Regression Methods

Kristína Machová, Marian Mach, Miroslava Hrešková

Department of Cybernetics and Artificial Intelligence, Technical University of Košice, Letná 9, 04200 Košice, Slovakia, kristina.machova@tuke.sk, marian.mach@tuke.sk, miroslava.hreskova@student.tuke.sk

Abstract: The paper is from the field of anti-social behavior recognition in online discussions. It focuses on the extraction of knowledge about special web reviewers as authorities or trolls, and trying to distinguish one from the other, based on their opinions and credibility. This paper presents a statistical application of supervised learning for creating a prediction model applied to find authorities (or trolls) among authors of online comments. The model can be used for differentiating between authoritative and non-authoritative reviewers. Standard methods, such as linear and logistic regression and genetic programming were applied. The objective of designed approach is to model dependency of the reviewer variable on independent predictors representing special reviewing. Values of those independent predictors (variables) are extracted from the data about structure as well as text content of online discussions. The model can offer important information for social web users, who search for truthful and reliable information, while in the process of developing their own opinions.

Keywords: reviewers classification; authority; troll; linear regression; logistic regression; genetic programming

1 Introduction

Currently, social media is studied intensively. Particularly, the extraction of summarized knowledge from online discourse content, has been a popular topic of research, in recent years. The volume of discourse data, extracted every day from social networks, is too large and time consuming to be processed by humans. Discourse data are accumulated from different online platforms, such as Facebook, Google+, Twitter, Disqus, Quora, etc. Within posts on social networks, useful information is mixed with misleading information, such as fake news, or troll reviews. Thus, it is important to search for authoritative sources and reviewers. An "authoritative-ness" evaluation becomes an increasingly important aspect of research on online discussion media. This evaluation can generate a list of reviewers, who are ordered according to their credibility. Usually, other users

follow the most credible reviewers. Our work can bring useful insights, for example, when we have to rely on the opinions given by reviewers, who are reviewing papers for conferences or journals.

Within this publication, we are focusing on the extraction of knowledge concerning specific reviewers of online communities. Our aim is to distinguish between classes of reviewers – authorities or trolls – automatically, in order to know whether we can trust their opinions. This approach is a part of research on the recognition of antisocial behavior in online communities. When classifying a reviewer into the authority or non-authority (troll) class, we have to address a classification problem. On the other hand, we can consider this problem as a regression problem, when we want to estimate the exact value of authoritativeness or trolling of some reviewer from a given interval. We have used the regression analysis, particularly linear, nonlinear, polynomial or symbolic regression as a measure of authoritativeness estimation, and logistic regression for web reviewer classification into the class of authority or non-authority.

2 Related Works

Our aim is to search for authorities and trolls of an online discussion. Other similar problems are authorship attribution, authorship verification of web reviewers and author profiling. Authorship attribution and identification is a problem of labelling an unknown document with a correct author from a given list of potential authors. The authorship attribution is based on measuring the similarity between the authors' interests (including writing styles) and the given text [1]. According to [2], character n-grams are considered among the best predictive features for the authorship attribution task. The paper presents an experiment with different machine-learning methods. The model for the authorship attribution can be successfully used for authorship verification as well. The objective of authorship verification [3] is to determine if a specific author has written a given text. The aim of the author profiling method is to learn all dimensions of the author profile. The paper [4] describes a method of learning an author profile with the focus on two dimensions: age and gender. The aim of this learning method is to identify differences in writing between a man and a woman within given dimensions.

A similar task was solved in work [5], which presents a methodology for detecting fake profiles on Twitter social network, and consequently, to associate them to real profiles in the same social network. This approach is based on an analysis of posts content generated by both fake and real profiles. Machine learning methods, such as Support Vector Machines, Random forests, K-NN and Naïve Bayes classifiers, were used. The results of accuracy were less than desirable. The highest achieved accuracy was 0.68.

The work [6] addresses problems related to detecting communities of authority and estimating the influence of such communities in dynamic social networks. The work detects communities sharing common interests – called "meta-communities", using topic modelling and the between-ness centrality. The authors empirically demonstrate the suitability of presented approach for the community-of-authority detection.

The approach [7] focuses on spreading of rumours in online discussions on Twitter. It is used to find an opinion leader using centrality measure metric on social network analysis. Aside from defining the centrality measure, it defines the special weighted edge. Twitter creates different kinds of relationship that can be turned into an edge, but not all the relationships have the same impact on spreading rumours. Thus, the study experiment considered edge weighting and centrality weighting. The study found that the edge with the ability to spread to a wider audience (quote, retweet, and reply) tends to have a bigger impact on finding an opinion leader. The study also finds that a low in-degree weight, high between-ness weight and low or no PageRank weight could give a 100% agreement upon other evaluation algorithms for finding the opinion leader.

Bouguessa et al. [8] proposes a parameter-less mixture model-based approach. They represent each user with a feature vector composed of information related to their social behavior and activity in an online community. Then, they propose a statistical framework, based on the multivariate beta mixtures, in order to model the estimated set of feature vectors. Therefore, the probability density function is estimated and the beta component which corresponds to the most authoritative users is identified.

The work [9] uses an assumption that the influencer has greater effect on the online social network than the average member. According to this definition, the influencer can be considered as an authority. In this work, an approach for influencer detection is designed, using semantic analysis to filter out irrelevant interactions, and achieving a simplified graph representation allowing the detection of true influencers. The approach is an interesting aggregation of both, semantic and social web.

According to paper [10], digital revolution has drastically changed people's lives. They try to distinguish the differences between two other user types - digital immigrants and natives. Digital natives are characterized by their highly automatic and quick response in a hyper-textual environment. Digital immigrants are characterized by their main focus on textual elements and a greater proneness to reflection. The main goal of the present research is to investigate the effect of affective priming on prosocial orientation in natives and immigrants by using a mobile application.

The majority of existing methods have certain limitations. There is usually a lack of automatic mechanism to formally differentiate between authoritative and non-

authoritative users. So we tried to find a formula for estimation of authoritativeness and trolling behavior of an online reviewer.

3 Authoritative Posting

The measure of "authority" can be represented by an ability of people to influence others, positively or negatively, and to lead them to achieve a certain goal. The concept of "authority" comes from the Latin word "augere". It denotes a person with opinions, attitudes or decisions respected by other group members. The authority can be derived from the relations between people (web users), but also from positions and hierarchies [11]. There are many kinds of authorities. Often, we divide them into *formal authorities* (measure of a person´s influence based on their formal position, regardless of personality traits) and an *informal* or *natural authorities* (result of personal capability, adequate self-confidence, spontaneous influence on others, no usage of pressure or force). Our work focuses on detection of informal – natural authority.

An important question to be answered is how the authority of social media can be defined. Reviewers insert their posts with their opinions, ideas and attitudes to an online discussion forum and thus create a "discourse content". The discourse content represents data for modelling an authority of online discussion, which is a special kind of authority. It is related not only to the content but also to the structure of an online discussion. We tried to find attributes (variables, predictors) representing the authoritative posting. The values of these attributes should be extracted from the structure of an online discussion as well as short texts of the posts. To achieve our objective, we need to define:

- Attributes (variables, predictors) – values of which can be extracted from an online discussion.
- Dependency of the variable "Authority" on the independent variables – attributes selected in the first step. This dependency should be represented by an estimation or discrimination function.

We will use capital letter A in "Authority" for a case, when Authority is not a general concept but an exact variable.

3.1 Attributes of Authority from the Structure

Each online discussion can be represented by a tree (Figure 1). The following information can be obtained from the discussion tree:

- *Number of discussion posts* by a given reviewer
- *Number of reactions* to posts(s) by a selected reviewer

- *Position in the discussion tree* expresses the mean level of posts by a reviewer
- *Word number* represents the ratio of the number of words in the reviewer's comments and the total number of words in a discussion

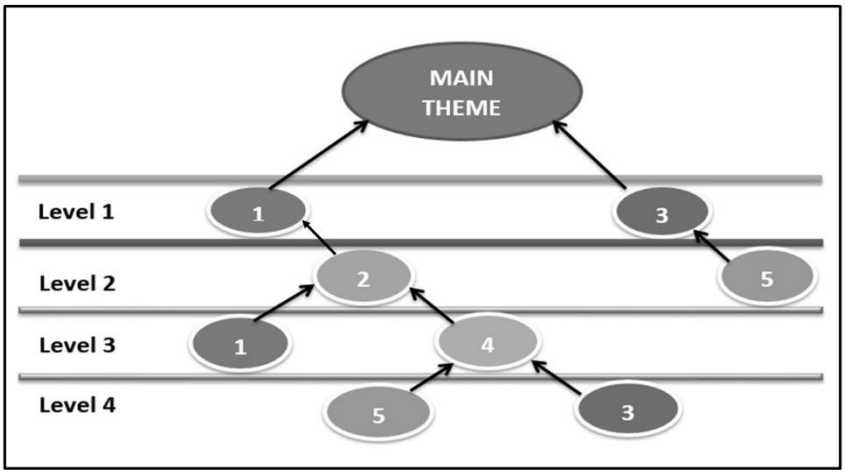

Figure 1
The discussion tree has 4 levels. The main theme is in the root. All comments of the same reviewer have the same number. Reactions are situated on levels 1 – 4

3.2 Attributes of Authority Related to Meta-information

Another approach to the authority attributes selection is based on their extraction of attributes from metadata about the reviewer posts in an online discussion. For example, evaluation of reviewer's authority by other reviewers:

- *Average evaluation of the comment* is represented by the ratio of the sum of all reactions (agree (+) and disagree (-)) on the posts of a given reviewer to the number of all their posts
- Value of *karma* is represented by a direct ratio of the number of readers who added a like to the total number of readers of the reviewer's posts. Karma is a number from the interval 0 to 200
- *Number of matched words* with a specific thesaurus
- *Number of likes* in one online discussion
- *Total sum of likes*
- *Number of followers*
- *Number of followed*

4 Used Regression Methods

We have used regression analysis methods [12] (linear, logistic and symbolic regression) to learn a model for the recognition of special online reviewers as authorities or trolls. The linear regression is based on formal definition of a line, which approximates the input data in the best way. It means that predicted values of depending variable (Authority in our case) are the most similar to the real values in the training data. The logistic regression is derived from the linear regression. These two methods vary in the type of variables used. The variable can be continuous or discrete.

- The *continuous variable* takes all values from a given interval of real numbers. For example, temperature from minus to plus values. Linear regression is a suitable method for modelling continuous variables. The continuous variables are used in solving the regression task – how to predict exact value of depended attribute for a new observation.
- The *discrete variable* takes values from a definite list of values. The values are categorical, for example man/woman or authority/non-authority. Logistic regression is a suitable method for modelling the discrete variable. Discrete variables are used in solving the classification task, i.e. how to predict the class for a new observation.

In short, the regression task represents an exact value prediction and classification task represents a class prediction.

The linear and logistic regression appeared to be a natural choice, because we wanted to obtain an estimation function, which can be linear because of the character of the data. We also wanted to learn the decision procedure for distinguishing an authoritative reviewer from a non-authoritative one based on the same data. Our aim was not only to learn the parameters of linear or other implicit functions but also to elicit the appropriate form of an estimation function without any conditions about its form. Therefore, we selected the genetic programming for this task. The genetic programming can not only learn an unexpected function, but can also randomly select only some of the variables (attributes) in training data to form an estimate of the function, thus providing useful information about the importance of authoritative posting attributes.

4.1 Linear Regression

The linear regression is a method for modelling dependency of variable Y (Y=Authority in our case) on the independent variable. It can be a simple Y=f(x) or a multiple regression, when we are modelling the dependency of the variable Y on several independent variables, called predictors, (x_1, x_2, …x_N) (1).

$$Y = f(x_1, x_2, \ldots, x_N) \tag{1}$$

The aim is to describe this relation by a suitable mathematical model, for example, by a linear function. The result will be a regression line, which should optimally match the empirical polygon [13]. The linear regression can be represented by the equation (2). Values of parameters (weights of predictors) $w_0, w_1, ..., w_N$ have to be found in order to achieve the optimal matching with the point graph (observations). These weights can be calculated using the Ordinary Least Squares method [14] for minimizing the sum of square mistakes.

$$y_i = w_0 + w_1 x_{i1} + ... + w_n x_{iN} + \varepsilon_i \tag{2}$$

4.2 Logistic Regression

When we use linear regression on data with a discrete variable, there appears to be a problem with the correct prediction. This problem (illustrated in Figure 2) arises when a sudden change (jump) occurs on the Y-axis – change from no reaction to reaction or similarly from non-authority to authority. Linear regression will incorrectly predict part of this change from value "response=0" to value "response=1". In this case, it is better to use a sigmoid curve obtained by the logistic regression. In Figure 3, we can see that the logistic regression is more suitable for predicting discrete variables.

If our aim is to estimate the value of reviewer´s authoritativeness from a given interval, the linear regression is a more suitable method. However, if we want to classify a web reviewer into the class of authority or non-authority, logistic regression is more suitable. Now the question is how to obtain the sigmoid curve, which is typical for logistic regression from the data. The logistic regression can be considered as an extension of linear regression. A line, i.e. graphical representation of linear regression, is defined by two constants: b_0 (cross point of the line) and b_1 (directional vector). In practice, the line can be used for prediction of value Y depending on X. Predicted value $Y' = b_0 + b_1 X$. In the case of logistic regression, the linear line has to be transformed to sigmoid curve. In such case, Y is defined as $Y=1$, if a reviewer is considered an authority, or $Y=0$, if a reviewer is not considered an authority. If p represents probability of the fact, that the given reviewer is an authority, then $(1-p)$ represents probability that the reviewer is not an authority and ratio of p and $1-p$ represents a chance that the reviewer is an authority. It is required for the logarithm of this chance to be linear for predictors X. This is modelled according to [15] by equation (3).

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 X \tag{3}$$

In the case of multiple independent variables $X_1, \ldots, X_N$, the logistic regression can be modelled using the equation (4):

$$\ln\left(\frac{p}{1-p}\right) = b_0 + b_1 X_1 + \cdots + b_n X_n \tag{4}$$

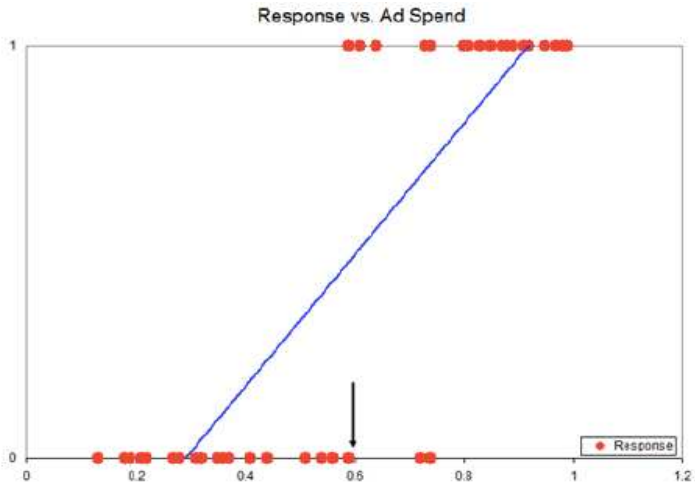

Figure 2

Linear regression in prediction of (no)reaction in dependence on an advertising spending [15]

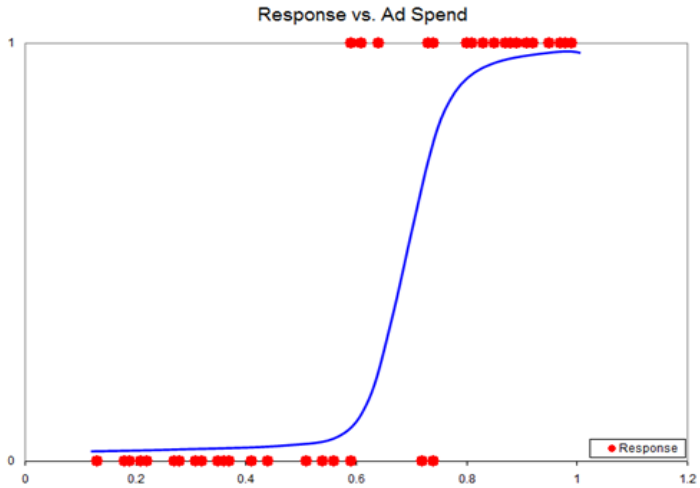

Figure 3

Logistic regression in prediction of (no)reaction in dependence on an advertising spending [15]

According to this model, the probability that a reviewer is an authority *(Y=1)* can be calculated according to equation (5):

$$p(Y=1) = \frac{e^{b_0+b_1X_1+\cdots+b_NX_N}}{1+e^{b_0+b_1X_1+\cdots+b_NX_N}} \tag{5}$$

After transformation, final model representing sigmoid curve has the following form (6):

$$p(Y = 1) = \frac{1}{1 + e^{-(b_0 + b_1 X_1 + \cdots + b_N X_N)}} \tag{6}$$

During the learning of this model, the values of constants (b_0, b_1, …, b_N) have to be learned from a training set.

5 Authority Modelling Using Classic Regression Analysis

5.1 Used Data

We have prepared a set of real data from an online discussion. We have used the same data set in all our experiments with different methods of statistical learning to compare the used techniques. The Authority value was related to reviewers, not to their comments. Therefore, we collected and aggregated all data about each reviewer. After the pre-processing step, the aggregated data had the form illustrated in Table 1. There were 117 reviewers acquired from the portal "www.sme.sk".

Table 1
Illustration of the data collected about each reviewer

Reviewer	AE	K	NCH	AL	ANR	NP
c1	60	108	26.0	**0**	1.0	1
c2	80	182	220.0	2	0.5	2
c3	80	171	548.5	3	2.5	2
c4	80	162	57.5	4	0.5	2
c5	50	99	112.5	6	0.0	2
…						

For each reviewer we collected values of the following independent variables – attributes of the web discussion:

- *NP* is the number of posts of a given reviewer
- *ANR* is the average number of reactions to the author's posts
- *AL* is the average number of layers, on which the posts of a reviewer are situated within the discussion tree (Figure 1)
- *NCH* is the average number of characters in a post by the reviewer
- *K* is the karma of a reviewer in the form of a number from 0 to 200, which represents the activity of a reviewer from the last 3 months

- *AE* is the average evaluation of the comment, which is available on the web discussion page. The values of *AE* are in the range from 0 to 80

The input data matrix was regular. There does not exist any strong correlation between any pair of variables – attributes. It is illustrated in Table 2.

Table 2
Correlation matrix of data from portal "www.sme.sk"

Attributes	**K**	**NCH**	**AL**	**ANR**	**NP**	**AE**
K	1	0.121	0.153	0.078	0.114	0.307
NCH	0.121	1	0.009	-0.024	0.162	0.348
AL	0.153	0.009	1	-0.027	0.354	0.015
ANR	0.078	-0.024	-0.027	1	0.071	-0.050
NP	0.114	0.162	0.354	0.071	1	0.190
AE	0.307	0.348	0.015	-0.050	0.190	1

In the Table 2, we can see values from the interval <+-0.01, +-0.29> representing a nearly zero correlation or values from the interval <+-0.30, +-0.49> representing a moderately medium correlation. There are no values representing a strong correlation (<+-0.50, +-1.00>). The types of correlations are presented in [16].

5.2 Linear and Non-linear Regression

The estimation function for the prediction of the value of authority can be represented by the formula (7):

$$Authority = f(NP, ANR, AL, NCH, K, AE) \quad (7)$$

At the beginning, we computed weights of linear and nonlinear regression functions. For computing these weights, we needed to know not only the values of independent variables *NP, ANR, AL, NCH, K, AE* but also the values of the dependent variable *Authority.* These values were obtained for each reviewer from:

- Labelling by a "human expert"
- Labelling by other reviewers of the online discussion – it represents the "wisdom of the crowd". This labelling represents a sum of information about a labelled example – reviewer, which was obtained from evaluations of this reviewer by other reviewers in the same discussion.

The following regression functions for authority estimation were generated: linear and non-linear functions learned from the "human expert" or "wisdom of the crowd". All these functions were generated using standard MATLAB functions: "regress" in the case of linear and "lsqnonlin" in the case of non-linear relations. No auxiliary regularization method was used because the input data matrix was regular. All learned functions have been tested using average deviation for the validation of estimation functions. After transforming the regression problem into a

classification problem, other tests were performed using classical precision and recall measures. The results can be found in Table 3.

Table 3
Achieved average deviation, precision and recall of tests of the approach to the authoritativeness identification based on linear and nonlinear regression

	DEVIATION		**PRECISION**		**RECALL**	
Version	EXPERT	CROWD	EXPERT	CROWD	EXPERT	CROWD
Linear	17.34	**3.29**	0.70	**0.98**	0.67	**0.80**
Non-linear	18.11	6.56	0.67	0.97	0.67	0.80

Once we have a suitable model for authority estimation, we can provide the information about an exact authority value of some reviewer from a defined interval to web users. The user might not need the exact value of authority but only an indication of whether the reviewer is or is not an authority. Thus, we tried to transform the regression problem to a classification problem by defining a threshold value for the variable Authority. All reviewers with the predicted value of Authority equal or higher than the defined threshold will be identified as authoritative reviewers. For this classification problem, the obvious measures precision and recall were used for the validation of classification to class Authority or Non-authority. For discrimination we used the *Threshold value of Authority = 70* from the interval (0, 100). The detailed results are in [17] and some of them are illustrated in Table 3 to compare them with results achieved by logistic regression and symbolic regression (genetic programming). The best results were achieved by learning a linear function from the "wisdom of the crowd" data.

We have also used RapidMiner Studio (version 7.4) for the creation of predictive models based on regression analysis. We chose this solution to avoid the need to generate an estimation function and use it in combination with the defined threshold to transform the regression problem to a classification problem. RapidMiner disposes of the operator "Classification by regression". The operator can use linear regression for prediction of a binary variable. But achieved results were very poor. That was the reason why we focused on a different solution using logistic regression, which is suitable just for the prediction of a categorical (binary) variable.

5.3 Logistic Regression

Besides our model based on linear regression, a model based on logistic regression was created using the operator "Logistic Regression" in RapidMiner. We set the solver to the value AUTO. We used "Split data" for transforming data into training and testing sets in the ratio of 70/30 percentage.

We performed experiments with the model based on linear and logistic regression using 10-fold cross validation. These experiments were done on three data sets:

www.sme.sk, www.quora.com, and www.disqus.com using the well-known measures Precision and Recall (see Table 4).

Table 4
Achieved Precision and Recall of tests of the designed approaches to authority classification based on linear and logistic regression

	www.sme.sk		www.quora.com		disqus.com	
Model	**Precision**	**Recall**	**Precision**	**Recall**	**Precision**	**Recall**
Linear regression	0.91	**0.97**	**0.97**	0.85	**0.86**	0.38
Logistic regression	**0.96**	0.95	0.94	**0.91**	0.80	**0.91**

From the results in Table 4, we can hardly say which is better, the model based on linear regression or the model based on logistic regression. Therefore, for easier comparison, we also present results of other measures such as Accuracy, F1 rate and AUC. Table 5 illustrates Accuracy and F_1 rate representing a harmonic average of Precision and Recall.

Table 5
Achieved Accuracy and F1 rate of tests of the designed approaches to authority classification based on linear and logistic regression

	www.sme.sk		www.quora.com		disqus.com	
Model	**Accuracy**	**F_1**	**Accuracy**	**F_1**	**Accuracy**	**F_1**
Linear regression	0.914	0.939	**0.929**	0.906	0.823	0.527
Logistic regression	0.914	**0.955**	0.917	**0.925**	**0.887**	**0.851**

The effectiveness of our models was verified by the means of an ROC curve [18] and an AUC value, which represents the area under the ROC curve.

- Values of AUC from the interval <0, 0.50) represent an inappropriate model
- AUC values from the interval <0.50, 0.75) represent an acceptable but not a very effective model
- AUC values from the interval <0.75, 0.92) represent a good model for the given data sample
- AUC values from the interval <0.92, 0.97) represent a very good model
- AUC values from the interval <0.97, 1.00) represent an excellent model for the given data sample

The results for linear and logistic regression are presented in Table 6. These results confirm that the model based on logistic regression is more suitable for Authority classification than the model based on linear regression.

Table 6
Achieved results of AUC values of tests of the designed approaches to authority classification based on linear and logistic regression

	www.sme.sk	www.quora.com	disqus.com
Model	AUC	AUC	AUC
Linear regression	0.50	0.50	0.50
Logistic regression	**0.98**	**0.95**	**0.96**

The results of the AUC values for linear regression are 0.50. The reason for this fact could be that linear regression is not primarily designed for the prediction of binary (categorical) attributes. Even the operator "Classification by regression" cannot improve the results of the model based on linear regression. On the other hand, the results of the AUC values for logistic regression evaluate this model as very good or excellent.

6 Genetic Programming as Symbolic Regression

Genetic programming is a modification of genetic algorithms which was designed by John Koza [19]. The chromosomes are represented by character strings in genetic algorithms. In genetic programming, these character strings are replaced by symbols, which can be represented by operators or by more difficult structures such as elementary functions. A chromosome, graphically represented as a syntactic tree, itself represents some function $y=t(x)$. This function assigns a dependent variable y to an independent variable x. For example, the function $t(x) = x(1+x)$ in Figure 4. Let $A= \{(x_i, y_i), i=1, ..., N\}$ be a training set containing N points. Then, our goal is to find the function t(x) by minimalizing the difference (quadratic or absolute) between the computed value of y and a real value of y from the training set. Particularly, when the sum of all the differences is smallest, we will have found the goal function represented by a syntactic tree. This formulation of the task is close to regression analysis, where only the parameters of a given function are learned for a training set of points. The space of functions is finite, and their form changes only based on changing parameters. In genetic programming the space of functions is infinite, so genetic programming represents a generalized approach to symbolic regression. The process of symbolic regression is based on the operations of crossover and mutation. Crossover is illustrated in Figure 5 and Figure 6. Two parents (syntactic trees) were selected as inputs for the operation of crossover. The special nodes emphasized with black circles are called "points of crossover" and they are selected randomly. After switching these points of crossover, we will obtain the new generation.

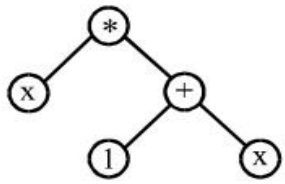

Figure 4

Syntactic tree as a representation of a chromosome corresponding to the function *x(1+x)* [20]

The process of symbolic regression is based on the operations of crossover and mutation. Crossover is illustrated in Figures 5 and 6. Two parents (syntactic trees) were selected as inputs for the operation of crossover. The special nodes emphasized with black circles are called "points of crossover" and they are selected randomly. After switching these points of crossover, we will obtain the new generation.

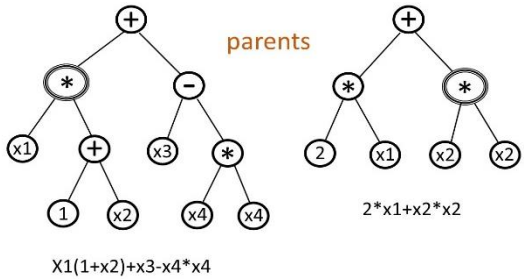

Figure 5

Illustration of parents' syntactic trees as an input to the operation crossover

The mutation operation (see Figure 7) comprises of replacing a subtree of the original tree with a new randomly generated subtree. This means that some branch of the syntactic tree is randomly changed. Besides these two elementary genetic operators, other secondary operators can be used, for example permutation, editing, encapsulation, decimation, etc. [20].

We need to use an individual in the form of a syntactic tree representing a function in genetic programming. The individual can be a computer program formed from functions and terminals. From these illustrations of crossover and mutation operations we can deduce that a random polynomial function can be constructed in the space of nodes with two arguments equivalent to the arithmetic operations of addition, subtraction and multiplication. In case we expand the set of acceptable nodes by those equivalent to new operations such as division, exponentiation etc., we can express practically any function. Work [21] contains more information about genetic programming.

Nowadays, genetic programming is often used for different classification problems. The work [22] offers a review of tree-based genetic programming

classification methods. It also analyses their strengths and weaknesses in relation to the task of classification.

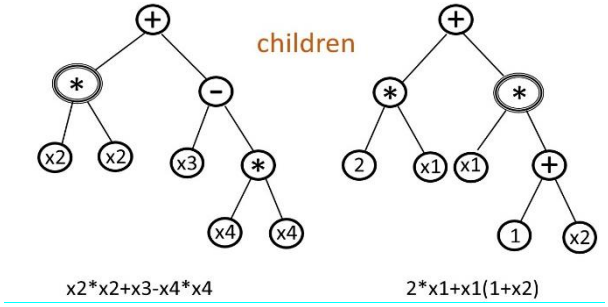

Figure 6

Illustration of children's' syntactic trees representing a new generation as the result of crossover over two syntactic trees in Figure 5

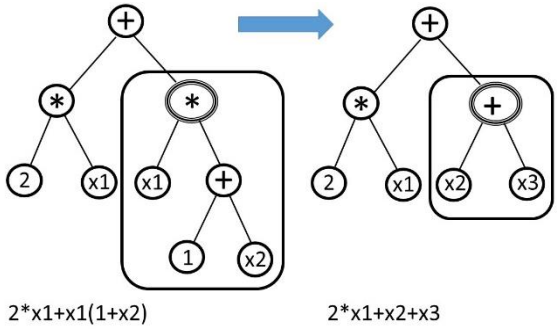

Figure 7

Illustration of mutation of one syntactic tree

In conclusion, genetic programming can be effectively used for performing the classifier evolution task. On the other hand, long training time and lack of convergence can be considered as its disadvantages. Nevertheless, it can be successfully used for this task.

7 Authority Modelling Using Genetic Programming

We have prepared an environment in Matlab for experiments with genetic programming to train a model for Authority estimation. At first we did some experiments with the finite although big space of elementary operators (functions). Nevertheless, the generated estimation functions were too complicated, completely unreasonable and it was impossible to interpret them. For this reason, we decided to constrain the space of elementary operators to the following:

addition, subtraction, multiplication and exponentiation to the second power. The function regfitness was used for the measure of fitness of individuals in the form: ind = regfitness (ind, params, data, terminals, varsvals).

We have used the data from section 6.1 to train the estimation function. The data were divided into three parts (123, 132 and 231) for three-fold cross validation (for example 123: 1st and 2nd parts for training and 3rd part for testing). Six experiments were conducted for both learning from an expert and from wisdom of the crowd. From the six functions generated in the process of learning using genetic programming, the function (8) learned from the "wisdoms of the crowd" on data 132 achieved the best results according Table 7.

$$A = plus(ANR,plus(NC,plus(NC,plus(AE,times(AL,NC))))) \quad (8)$$

All learned functions have been tested using classical Precision, Recall and F1 measures. The results can be found in Table 7. The results are promising (between 94 and 100 percent), with the exception of Precision of the models trained on expert labelled data (in average 59 percent) and consequently a lower F1 rate. Overall results for all used methods of regression analysis: linear, non-linear, logistic and symbolic (genetic programming) are presented in Table 8.

Table 7

Achieved results of Precision, Recall and F1 rate of tests of the approach based on genetic programming

	Precision		**Recall**		**F1**	
Data	EXPERT	CROWD	EXPERT	CROWD	EXPERT	CROWD
123	0.592	0.937	0.938	1.000	0.726	0.967
132	**0.623**	**0.974**	**0.979**	**1.000**	**0.762**	**0.987**
321	0.562	0.961	0.911	1.000	0.695	0.980
Average	**0.592**	**0.957**	**0.943**	**1.000**	**0.728**	**0.978**

According to the values of the F1 measure (a balanced averaged of Precision and Recall) in Table 8, the best model for authority classification is the model trained by symbolic regression (genetic programming) on data labelled by a crowd. Using logistic regression, we have trained the second best model on the same data. In the case when we only consider the precision measure reflecting the number of false positive classifications, the best model is that learned using linear regression on crowd data. On the other hand, if we only consider the recall measure reflecting the number of false negative classifications, the best model is the one trained using symbolic and then logistic regression on crowd data.

The genetic algorithms used have brought new information about a real need and suitability of independent attributes of authoritative posting. Table 9 presents the frequency of appearance of these attributes (AE, ANR, AL, NC, NCH and K) in 16 experiments within the space of elementary operators (functions) and 6 experiments with a constrained space of elementary functions in the process of generation of the Authority estimation functions.

Table 8
Achieved results of Precision, Recall and F1 rate of tests in all used methods on the same data

	Precision		**Recall**		**F1**	
Regression	EXPERT	CROWD	EXPERT	CROWD	EXPERT	CROWD
Linear	0.70	**0.98**	0.67	0.80	0.68	0.88
Non/linear	0.67	0.97	0.67	0.80	0.67	0.87
Logistic	-	0.91	-	**0.97**	-	**0.94**
Symbolic	0.59	**0.96**	**0.94**	**1.00**	0.73	**0.98**

Table 9
Frequency of appearance of the monitored attributes in generated functions using genetic programming within 22 experiments

Attribute	**Frequency**	**%**
AE – average evaluation	22	100.00
ANR – average number of reactions	14	63.64
AL – average layer	13	59.09
NC – number of comments	11	50.00
NCH – number of characters	7	31.82
K – karma	7	31.82

Conclusions

The contribution of this paper is the design of a novel approach to the recognition of authority or trolling for special online reviewers. This approach is based on the solution of the problem of Authority identification from a discourse content. A similar solution could be used for Troll classification when attributes of troll posting are defined. The first solution was based on learning the Authority estimation function from the structure of online discussions using linear or non-linear regression. According to the test results, the best solution was a linear regression function trained from "wisdom of the crowd". The second solution was the classification model based on logistic regression. According to the test, the logistic regression method is more suitable for the identification of special web reviewers as an Authority. The last solution is a model obtained by symbolic regression in the form of genetic programming. This solution seems to be the most suitable. The results in Table 7 and Table 8 also prove, that learning from data labeled by crowd is more precise, no matter which learning method was used.

Authority identification can be used in many real situations. Mostly when web users search for an authority that can advise them in decision-making or forming an opinion on important topics. The work can have many useful applications, especially in areas where we have to rely on the opinions of reviewers, for example the case of reviewing papers for conferences or for journals.

The decision making process can be influenced by trolling and other types of antisocial online reviewing. For the future, we would like to focus on the

modelling of troll reviewers in online discussions within the field of antisocial behavior recognition.

When we talk about sentiment analysis, we usually mean the analysis of opinions and emotions. Emotion analysis can be used in the field of robotics in human-robot interaction. Particularly emotion analysis of a human can be useful for a robot, which communicates with humans to be able to adapt its behavior to the emotional situation of the interaction.

Other future extensions are connected to the semantically enriched detection of special web reviewers using an ontology [23] and involving topic modeling [24] of the texts of the examined reviews. It would also be useful to involve a neural network as a widely used method, in extended testing. We may only use the variables AE, ANR, AL and NC for the generation of the Authority estimation function. A surprising insight was that the "*karma*" of a reviewer doesn't seem to be very influential in determining his/her authority.

Acknowledgement

This work was supported by the Slovak Research and Development Agency under No. APVV-017-0267 "Automated Recognition of Antisocial Behavior in Online Communities", the contract No. APVV-016-0213 and APVV-015-0731.

References

[1] Yang, M., Cui, T. Tu, W.: Ordering-sensitive and Semantic-aware Topic Modelling. In: Twenty-Ninth AAAI Conference on Artificial Intelligence, 2015

[2] Markov, I., Baptista, J., Pichardo-Lagunas, O.: Authorship Attribution in Portuguese Using Character N-grams. Acta Polytechnica Hungarica, Vol. 14, No. 3, 2017, ISSN 1785-8860, 59-78

[3] Stamatatos, E., Fakotakis, N., Kokkinakis, G.: Automatic Text Categorization in Terms of Genre and Author, Computational linguistics 26, 2000, 471-495

[4] Mechti, S. Jaoua, M., Belguith, L. H., Faiz, R.: Author Profiling Using Style-based Features. In: Notebook for PAN at CLEF, 2013, 1-7

[5] Galán-Garcia, P., De La Puerta, J. G., Gómez, C. L., Santos, I., Bringas, P. G.: Supervised Machine Learning for the Detection of Troll Profiles in Twitter Social Network: Application to a Real Case of Cyberbullying. Logic Journal of the IGPL, Vol. 24, No. 1, 2014, ISSN 1367-0751, 42-53

[6] Chikhaoui, B., Chiazzaro, M., Wang, S., Sotir, M.: Detecting Communities of Authority and Analyzing Their Influence in Dynamic Social Networks. ACM Transactions on Intelligent Systems and Technology, Vol. 8, No. 6, 2017, article 82, ISSN 2157-6904, 1-28

[7] Devi, FK., Yudhoatmojo, SB., Budi, I.: Identification of Opinion Leader on Rumor Spreading in Online Social Network Twitter Using Edge Weighting and Centrality Measure Weighting. In. Proc. of the 12th International Conference on Digital Information Management (ICDIM), Kyushu Univ, Fukuoka, JAPAN, SEP 12-14, 2017, 313-318

[8] Bouguessa, M., Ben Romdhane, L.: Identifying Authorities in Online Communities. ACM Transactions on Intelligent Systems and Technology, Vol.6, No.3, article 30, 2015, ISSN 2157-6904, 1-23

[9] Ríos, SA., Aguilera, F., Nuñez-Gonzalez, JD., Graña, M.: Semantically Enhanced Network Analysis for Influencer Identification in Online Social Networks. NEUROCOMPUTING, 2017

[10] D'Errico, F., Paciello, M., Fida, R., Tramontano, C.: Effect of Affective Priming on Prosocial Orientation through Mobile Application: Differences between Digital Immigrants and Natives. Acta Polytechnica Hungarica, Vol. 16, No. 2, 2019, 109-128

[11] Chavalkova, K.: Authority of a Teacher (in Czech). Philosophic faculty of the University of Pardubice, Pardubice, Czech republic, 2011

[12] Chatterjee, S., Simonoff, J.S.: Handbook of Regression Analysis. A John Wiley & Sons, INC. publication, 2013, 1-236

[13] Pazman, A., Lacko V.: Lectures from Regression Models (in Slovak). University of Comenius Bratislava, Bratislava, Slovakia, 2012, ISBN 978-80-223-3070-1, 132 ps.

[14] Pohlman, J. T., Leitner, D. W.: A Comparison of Ordinary Least Squares and Logic Regression. The Ohio Journal of Science, Vol. 103, No. 5, 2003, 118-125

[15] Deshpande, B.: Logistic Regression Digest. http://resources.simafore.com/hs-fs/hub/64283/file-15121709-pdf/pdfs/Simafire_logistic_regression_article_digest.pdf?t=1493427943502, Accessed 29 November 2019

[16] Zlacká, A.: Correlation dependence. http://www.fhpv.unipo.sk/cvt/statistika/zlacka/geoinfo7.pdf, 2019, Accessed 29 November 2019 (in Slovak)

[17] Machová, K., Štefaník, J.: Regression Methods in the Authority Identification within Web Discussions. In: Computational Collective Intelligence, LNAI, Vol. LNAI 9329, No. 1, 2015, Springer-Verlag, Berlin Heidelberg, ISSN 0302-9743, 203-212

[18] Bortlíček, Z.: ROC Curves. Masarik University of Brno, Science Faculty, Brno, 2018, 1-39

[19] Koza, J. R.: Genetic Programming: On the Programming of Computers by means of Natural Selection. MIT Press, Cambridge, MA, 1992

[20] Kvasnička, V., Pospíchal, J., Tiňo, P.: Evolutionary Algorithms (5. Genetic Programming). (in Slovak) Publishing House of STU, Bratislava, 2000, 1-215

[21] Návrat, P.: Artificial Intelligence: Metaheuristics and Evolutionary Algorithms. (in Slovak) Publishing House of STU, FIIT STU, Bratislava, 2014, 1-418

[22] Jabeen, H., Baig, A.R.: Review of Classification Using Genetic Programming. International Journal of Engineering Science and Technology, Vol. 2, No. 2(2010), ISSN 0975-5462, 94-103

[23] Machová, K., Vrana, J., Mach, M., Sinčák, P.: Ontology Evaluation Based on the Visualization Methods, Context and Summaries. Acta Polytechnica Hungarica, Vol. 13, No. 4, 2016, 53-76

[24] Smatana, M., Butka, P.: TopicAE: A Topic Modeling Autoencoder. Acta Polytechnica Hungarica, Vol. 16, No. 4, 2019, 67-86