A Novel Entropy-Based Sensitivity Analysis Approach for Complex Systems

Ingrid Kovacs[1], Alexandra Iosub[2], Marina Țopa[1]
[1]Technical University of Cluj-Napoca, Romania
[2]Gheorghe Asachi Technical University of Iași, Romania
ingrid.kovacs@staff.utcluj.ro

Andi Buzo[3], Georg Pelz[3]
[3]Infineon Technologies AG, Neubiberg, Germany

Abstract— **The modern electronic systems have become very complex with a high number of potential factors that may affect the systems' behavior. Sensitivity analysis may be employed to simplify the analysis of such systems and identify the most important factors upfront. The paper introduces two new sensitivity analysis methods based on the measure of entropy, which overcome the limitation of several state-of-the-art methods imposing a specific design of experiments and a high computational cost, measured as the number of simulations (measurements) needed, for the sensitivity analysis. Their performance is compared to other methods based on variance decomposition and One-Factor-at-a-Time screening. The proposed methods named the Entropy Simple method and the Entropy Pair one are applied on a set of custom functions and an E-Bike application. They proved to have comparable accuracy to the state-of-the-art methods with the advantage of a lower computational cost and which does not increase with the number of factors.**

Keywords— ***Sensitivity analysis, entropy, variance decomposition***

I. INTRODUCTION

Modern electronic technology brings systems of high complexity, which leads to the need of new analysis techniques. The responses of these systems may be influenced by hundreds of input parameters, called factors. Examples of such factors are configuration settings, design parameters, stimuli variation, etc. The verification of such complex systems is time-consuming and costly. In order to shorten the verification time, one could reduce the analysis to a smaller number of factors of the system; it is often the case that only a few factors influence the variability of the system response. Examples of system responses are power consumption, rising time, noise level, etc.

The problem of determining the subset of factors most influential on the response may be addressed by sensitivity analysis (SA) [1-2]. Common statistical procedures for SA are Analysis of Variance (ANOVA) [3], statistical regression [4-5] and the correlation coefficient method [6]. However, for systems with a large number of factors or nonlinear factor-response relationships, these methods have poor accuracy. The standardized regression coefficient (SRC) method [4] can be applied only for linear factor-response relationships, otherwise the results are difficult to explain. Visual methods, such as the matrix scatter plots [3] are able to detect complex factor-response relationships, but their disadvantage is that they require human intervention for judgment on the factors' impact, which for a high number of factors becomes intractable.

SA may be employed also for uncertainty and worst-case-analysis. The choice of a SA method depends on the objective of the research, as well as on the types of factors involved and the execution cost implied. Here, by execution cost we mean the number of system evaluations, i.e. simulations or measurements.

SA methods were successfully applied in domains, e.g. atmospheric chemistry [7], transport emission [8] or fish population dynamics [9]. For electronic systems, however, the common SA methods applied are the One-Factor-at-a-Time (OFAT) method [10] and the partial derivatives [11], which are local methods and do not explore the full input space.

In this paper, we propose two efficient SA methods based on the measure of entropy. We also provide a comparative assessment of their accuracy to several SA methods from the literature based on variance decomposition [12-13] and the Morris OAT screening [13]. The comparison is performed on a set of custom-defined test functions and an E-Bike application. The SA approaches are compared from the point of view of factor ranking and the execution cost implied.

The paper is structured as follows: Section II describes the theoretical framework of the proposed entropy-based SA methods, followed by a summary of the applications along with the obtained results on the custom test functions and on the E-Bike application in Section III and Section IV, respectively. Conclusions are drawn in Section V.

II. PROPOSED SA METHODS

Using the measure of the statistical entropy, two SA methods were developed: the Entropy Simple method, which is capable of determining main effects of factors and the Entropy Pair method, which can determine also interactions between factors.

The core idea is based on the intuition given by scatter plots. For simplification, let us illustrate the idea through a simple example. Let (1) describe an example of factor-response relationship, where $F_1 \div F_4$ are the factors and Y denotes the response. Equation (1) includes a main effect of F_2, a quadratic effect of F_1 and an interaction effect between F_2 and F_3; F_4 has no effect on Y.

$$Y = 8F_1^2 + F_2 - 12F_2F_3 \tag{1}$$

The factor-response relationship is illustrated by the scatter plots in Fig 1. Note that F_4 has no impact on Y and this fact is revealed by the random distribution in the F_4-Y quadrant of Fig. 1, while for all other factor-response distributions one can observe an order of the distribution (or a deformation from randomness).

Based on the above observations, the principle of the entropy-based methods is the following: if the distribution of a response is independent of the distribution of the factor, it means that the factor has no impact on the response (see F_4-Y quadrant of Fig. 1).

Furthermore, if we divide the factor-response distribution into subsets with respect to the response value (e.g. subset 1: response < median and subset 2: response > median) and compare the distribution of the subset to the total distribution of the data, we can determine the impact of the factor on the response. The same principle can be applied on pairs of factors, leading to the ability of determining interactions between factors.

The term entropy and its formula are borrowed from literature and is used here as a measure of the randomness of a system [14]. Note that lower entropy means greater impact of the factors on the response. Shannon defined the statistical entropy in [14] as:

$$H(Y) = -\sum p_i \ln(p_i) \tag{2}$$

where Y has a discrete distribution and $p_i = P(Y{=}Y_i)$. We used the measure of entropy as an objective measure of the randomness of a system.

Fig. 2 describes the steps of the Entropy Simple SA approach in detail. The first step is similar to any SA approach, where the experiment is planned and the responses are measured. Note that if there are simulations/measurements available from other types of analysis, this step can be skipped, because opposite to the state-of-the-art SA methods, the entropy-based methods do not impose a specific experiment plan for the SA. Otherwise, we recommend a Monte Carlo (MC) design where the factors take random values uniformly distributed from their allowed interval. This assures a better coverage of the factors' space.

In the second step, the factor-response space,(F_i, Y), is divided into rectangles, also called bins. The edges of each bin are the quantiles of distributions of Y and F_i. For example, if we decide to divide the space 10 by 10, the edges of the bins will be the 10-quantiles of distribution Y and 10-quantiles of distribution F_i.

Then, the number of data points m_b falling in each bin b from a total of B bins, are counted. Their sum is equal to the total number of runs, N:

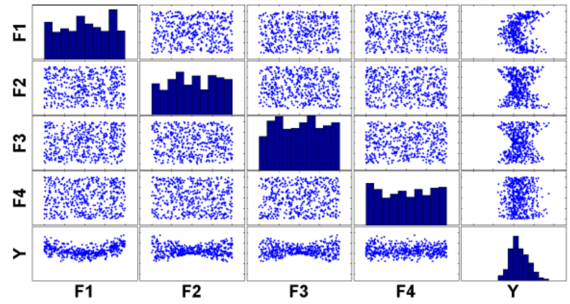

Fig. 1. The scatter plot of the factor-response relationship of (1)

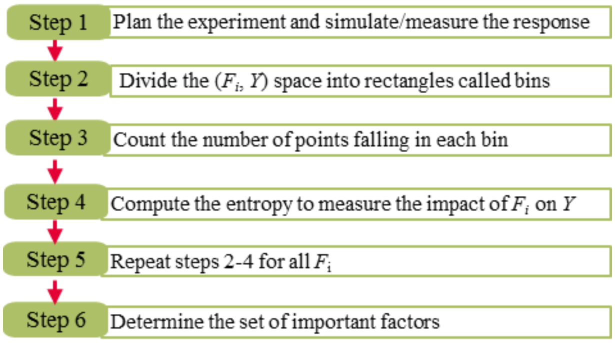

Fig. 2. The steps of the Entropy Simple SA method

$$N = \sum_{b=1}^{B} m_b \tag{3}$$

Note that the number of bins is arbitrarily chosen and depends on the number of observations. From statistical point of view, a bin should be large enough such as to cointain sufficient data points in average (>20).

In step 4, the entropy of the space formed by F_i and Y is computed as:

$$H_i = -\sum_{b=1}^{B} \frac{m_b}{N} \log\left(\frac{m_b}{N}\right) \tag{4}$$

H_i is computed for each F_i and in the end, the most important factors are determined. Note that the factor with the lowest H_i has the greatest impact on the response.

The approach for the Entropy Pair SA method is similar, but here, 3D bins of (F_i, F_j, Y) are formed and they will be rectangle prisms with edges the quantiles of F_i, F_j and Y.

Also, each H_{ij} has to be computed, which will lead to a KxK symmetric matrix, where K denotes the total number of factors and H_{ij} represents the entropy of (F_i, F_j, Y). Note that $H_{ij}=H_{ji}$. Again, the pair with the lowest entropy has the greatest impact on the response. Then, the impact of each individual F_i on Y is computed as:

$$H_{Ti} = \min_i(H_{ij}) \tag{5}$$

where H_{Ti} measures the total impact of factor F_i on Y.

Note that H_i and H_{Ti} are metrics which aim to give information about the main and total effect contribution of each factor on the response, similar to S_i and S_{Ti} metrics of the variance-based SA methods [15]. The common judgement of sensitivity indices, including S_i and S_{Ti}, is that a higher index means higher impact. Therefore, we rescaled H_i and H_{Ti} such that a higher H_i or H_{Ti} means higher impact. For example, we obtained the following H_{Ti} indices for (1): H_{T1}=0.316, H_{T2}=1, H_{T3}=1, H_{T4}=0.

One advantage of these methods over the methods from literature is that they do not impose a particular design of experiments in order to perform the SA, meaning that they can be performed with a design of experiments created for another type of analysis (e.g. pass/fail, yield analysis, etc). This is possible because the range of a factor or a response is not divided into equal intervals, but into quantiles which can be calculated regardless the type of distribution. Also, the methods do not have any restriction about the orthogonality of the factors, which means that correlations between factors may be induced.

III. Evaluation of the SA Methods on Custom Test Functions

First, the methods were tested via a mathematical example where the relations are known, in order to allow a comparison between the methods' prediction and the experimenter's expectation. The custom defined test functions are polynomial functions of the form of (6), which included different types of factor effects (main, quadratic and first order interactions) and the coefficients β_i and β_{ij} measure the importance of factor x_i:

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \sum_{j \geq i}^{k} \beta_{ij} x_i x_j \qquad (6)$$

A set of 60 polynomial functions of the form of (6) was implemented, each including 30 factors. For each function, we selected a set of important (target) factors. To test the robustness of the methods we added noise on the response as in (7):

$$y_{i\,noise} = y_i + \sigma_{noise} \cdot \eta_i \qquad (7)$$

where η_i are normally distributed pseudorandom numbers. Noise scenarios represent the case where the analysis is made on measurement data and not simulation ones.

In order to quantify the effect of the noise, the measure of Signal-to-Noise Ratio (SNR) was considered and it was defined as the ratio of the variance of the response and the variance of the noise:

$$SNR_y = \frac{\sigma_y^2}{\sigma_{noise}^2} \qquad (8)$$

The methods were tested with seven SNR values SNR_{dB} ={25, 20, 15, 10, 5, 0, -5} dB. At SNR=25 dB the same accuracy was obtained for all methods as if no noise was added.

We defined the *pass rate* as the accuracy testing criteria of the methods. This measure determines the percentage of selected important (target) factors, x_t, which have also been returned in the top of the five most important factors of a method.

As the polynomial functions included 2-4 important factors, we found it a sufficient condition that the factor is returned in the top five most important factors.

Fig. 3 illustrates a comparison of the accuracy of the proposed methods to several variance-based methods and the Morris method for noise added on y. The FAST method is capable of detecting only the main effects and it was tested only for this type of effect. The proposed entropy-based SA methods outperformed the Morris and Sobol methods and had similar accuracy with the EFAST and Jansen methods.

From the point of view of the execution cost, meaning the number of simulations required for the SA analysis, the proposed methods were the most efficient, as they were performed on a uniform Monte Carlo design with 300 runs, followed by the Morris method which required 310 runs. The high accuracy of the Jansen and EFAST methods implies a high computational cost, because they required 900 and 1950 runs, respectively. The FAST method required 23081 runs, while the Sobol method implied 1830 runs.

This study provided valuable information about the accuracy of the proposed methods compared to the other SA methods, highlighting the advantage of a lower execution cost and it served as reference for the study of a real system application, where there is little information about the effects of each factor on the output.

IV. Evaluation of the SA Methods on the E-Bike Application

The knowledge provided by the tests on the synthetic functions served as reference for the study of a real system, which was an E-Bike application model [16-17]. An E-Bike is a regular bicycle with an integrated electric motor to provide additional assistance. The analysis included 15 factors (see in Table I) and two responses (see in Table II). Note that the factors are of different nature (system architecture properties, component properties, system inputs, operating conditions).

For the SA we selected the proposed entropy-based methods and the EFAST method, because the latter gained very good accuracy on the synthetic functions and served as reference of the true factor-response sensitivities. The factors-responses scatterplots revealed nonlinear relationships between certain factors and responses, this is why simple methods such as SRC were not considered.

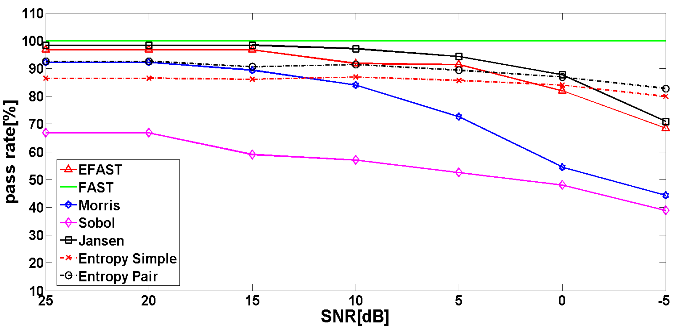

Fig.3. Accuracy comparison of the SA methods; noise added on y

The purpose of the experiments was to assess and compare the accuracy of each method on an application. We applied each method for both responses and the results were compared in terms of the returned ranking of the factors, by considering the ranking of the EFAST method as reference.

The high accuracy of the EFAST method (proved previously on the synthetic functions) implied the compromise of a high computational cost which translates also into a high computational time. The simulations are costly because the implemented E-Bike uses a direct driven motor with a large mechanical time constant. Given the imposed computational cost as a function of the factors implied [15], the EFAST method required 975 simulations for the SA approach, which translates into a computational time of approximately 112 hours.

First, we performed the simulations according to the design of experiments imposed by the EFAST method. Then, we applied the SA for both the EFAST and entropy-based methods, using the advantage of the entropy-based methods that they do not impose a particular design of experiments and that they can be applied on simulations resulted from other analysis.

For validation of the results, we compared the top of the most important factors returned by each method, considering the true top of important factors the one returned by the EFAST method.

A subsequent analysis was to apply the entropy-based methods on a lower number of simulations and to compare the returned top of the most important factors to the top of the EFAST method. The purpose of this analysis was to determine if the entropy-based methods determine the same important factors even with a lower number of simulations.

As the system contained a number of 15 factors, we considered the top five for the *AccTime* response, respectively top four for the *TorqueRipple* of the most important factors, as usually only a lower number of factors have impact on the output response.

Table III illustrates the results of the SA analysis with the entropy-based methods performed on the simulation results of the EFAST method. The top five most important factors of the Entropy Simple and Entropy Pair methods is the same as the top of the EFAST method. Also, the ranking of the first three factors, i.e. *HumanInertia*, *Ke* and *GainA,* is the same as for the EFAST method, while the ranking of factors *Rshunt* and *Wref* is interchanged compared to the EFAST method. However, *Rshunt* and *Wref* have approximately the same total sensitivity index, S_{Ti}, and hence similar impact on the *AccTime* response, which explains the swapping of the two factors by the entropy-based methods.

One subsequent analysis was to apply the entropy-based methods on a lower number of simulations and test if the top of the most important factors is similar to the EFAST method, performed on 975 simulations. For this we used a uniform Monte Carlo design with 300 runs, as this is the common number of simulations for a Monte Carlo experimental setup.

TABLE I. THE FACTORS OF THE E-BIKE APPLICATION

Components	Factors Label [units]	Description
Current Sensor	*GainA* [-], *OffsetA* [μV]	Gain of the sensor amplifier; Offset voltage of the sensor amplifier
	LevelNoise [-]	Adjustment factor for the level of noise floor of the amplifier
	RoLPF [Ω] *CoLPF* [nF]	Resistance and capacitance of the output Low Pass Filter
	Rshunt [$\mu\Omega$]	Shunt resistor
Angle Sensor	*OffSin* [V] *ASin* [-] *PhiY* [ ]	Offset in the sine sensor voltage; Synchronicity error; Mechanical misalignment
Motor	*Rs* [Ω] *Ls* [H]	Motor resistance; Motor inductance
	Ke [V/rad/s]	BackEMF voltage constant
Operating Conditions	*Wref* [rad/s]	Speed reference
	HumanInertia [kgm^2]	Human moment of inertia
Inverter	Inver*terSupply* [V]	Supply of the inverter

TABLE II. RESPONSES NAMES AND SIGNIFICANCE

Response name	Significance
AccTime	Acceleration time
TorqueRipple	Torque ripple

Table IV illustrates the results. Even with a computational cost three times smaller than for the EFAST method, the entropy-based methods identified four factors from a total of five important factors in the top of the most important factors. *HumanInertia* and *Ke*, which were returned to have the greatest impact on the *AccTime* response by the EFAST method, are ranked in the top of the three most important factors according to the entropy-based methods.

The same approach was repeated also for the *TorqueRipple* response. Table V presents the results of the SA with the entropy-based methods applied on the simulations of the EFAST experimental setup. Three factors from the top of the four most important factors of the EFAST are returned as important factors also by the entropy-based methods. Moreover, the *Wref* and *LevelNoise* factors are ranked similarly by all methods.

The same approach was applied on a lower number of simulations, considering a uniform Monte Carlo design with 300 runs. Table VI illustrates the results. The ranking of the important factors according to the Entropy Simple method remains the same even with 300 simulations, while the ranking of the top three most important factors of the Entropy Pair method corresponds to the top three of the EFAST method.

The proposed entropy-based methods proved their efficiency in terms of the computational cost and also accuracy, having approximately the same accuracy on 300 simulations as the EFAST method on 975 simulations, which means a 60% lower computational cost than the EFAST method.

TABLE III. TOP OF MOST IMPORTANT FACTORS; *ACCTIME* RESPONSE; 975 SIMULATIONS FOR THE ENTROPY-BASED METHODS

EFAST (975 runs)		**Entropy Simple (975 runs)**		**Entropy Pair (975 runs)**	
factor	S_{Ti}	**factor**	H_i	**factor**	H_{Ti}
HumanInertia	0.629	*HumanInertia*	1	*HumanInertia*	1
Ke	0.228	*Ke*	0.149	*Ke*	1
GainA	0.073	*GainA*	0.113	*GainA*	0.767
Rshunt	0.035	*Wref*	0.102	*Wref*	0.608
Wref	0.033	*Rshunt*	0.042	*Rshunt*	0.330

TABLE IV. TOP OF MOST IMPORTANT FACTORS; *ACCTIME* RESPONSE; 300 SIMULATIONS FOR THE ENTROPY-BASED METHODS

EFAST (975 runs)		**Entropy Simple (300 runs)**		**Entropy Pair (300 runs)**	
factor	S_{Ti}	**factor**	H_i	**factor**	H_{Ti}
HumanInertia	0.629	*HumanInertia*	1	*HumanInertia*	1
Ke	0.228	*Wref*	0.148	*GainA*	1
GainA	0.073	*Ke*	0.115	*Ke*	0.770
Rshunt	0.035	*ASin*	0.097	*Wref*	0.548
Wref	0.033	*GainA*	0.065	*LevelNoise*	0.340

TABLE V. TOP OF MOST IMPORTANT FACTORS; *TORQUERIPPLE* RESPONSE; 975 SIMULATIONS FOR THE ENTROPY-BASED METHODS

EFAST (975 runs)		**Entropy Simple (975 runs)**		**Entropy Pair (975 runs)**	
factor	S_{Ti}	**factor**	H_i	**factor**	H_{Ti}
Wref	0.426	*Wref*	1	*Wref*	1
LevelNoise	0.318	*LevelNoise*	0.836	*InverterSupply*	1
Rshunt	0.249	*PhiY*	0.404	*LevelNoise*	0.767
InverterSupply	0.248	*Rshunt*	0.303	*ASin*	0.417

TABLE VI. TOP OF MOST IMPORTANT FACTORS; *TORQUERIPPLE* RESPONSE; 300 SIMULATIONS FOR THE ENTROPY-BASED METHODS

EFAST (975 runs)		**Entropy Simple (300 runs)**		**Entropy Pair (300 runs)**	
factor	S_{Ti}	**factor**	H_i	**factor**	H_{Ti}
Wref	0.426	*Wref*	1	*Wref*	1
LevelNoise	0.318	*LevelNoise*	0.661	*LevelNoise*	1
Rshunt	0.249	*PhiY*	0.253	*Rshunt*	0.701
InverterSupply	0.248	*Rshunt*	0.240	*ASin*	0.625

Moreover, the proposed methods have the advantage that they do not impose a particular design of experiments and could be applied also on the simulation results of the EFAST experiment plan.

V. CONCLUSIONS

Two novel entropy-based methods were introduced and a comparison of their performance in terms of accuracy and computational cost to the state-of-the-art methods was done for two setups. First, the methods were applied on a set of custom-defined functions and then an E-Bike application, both including a high number of factors.

The reason for using the test functions was to do a comparison in a controlled manner of the methods' performance in terms of factor ranking for different types of effects and to determine the execution cost implied by each method.

On the test functions, the proposed methods outperformed the Morris and Sobol variance-based methods and had similar accuracy as the Jansen and EFAST methods, with the advantage of lower computational cost (which does not increase with the number of factors).

The methods were also evaluated on a real life application, an E-Bike. In this case, the results of the EFAST method were considered as reference (because of their high accuracy on the test functions) and we compared the rankings provided by the proposed entropy-based methods to the ranking of the EFAST method.

As the proposed entropy-based methods have no restrictions about the experimental setup nor the number of simulations required, we applied them on the experimental runs resulted from the EFAST SA approach (975 runs) and on a uniform Monte Carlo design with 300 runs.

Even with a much lower computational cost, the proposed methods returned the factors labeled as important by the EFAST method. By such an analysis we were able to determine the most influential factors on the E-Bike's responses of interest.

ACKNOWLEDGMENT

This work was co-funded by the European Regional Development Fund through the Operational Program "Competitiveness" POC -A1.2.3-G-2015, project P_40_437, contract 19/01.09.2016.

REFERENCES

[1] A. Saltelli, S. Tarantola, F. Campolongo, M. Ratto, "Sensitivity analysis in practice: A guide to assessing scientific models", John Wiley & Sons Ltd., Chichester, 2004.

[2] S. Tarantola, D. Gatelli, T.A. Mara, "Random balance designs for the estimation of first order global sensitivity indices", Reliability Engineering and System Safety, vol. 91, pp. 707-717, 2006.

[3] G. W. Snedecor, W. G. Cochran, "Statistical Methods", 8th ed., Iowa State University Press, 1989.

[4] D. Montgomery, "Design and Analysis of Experiments", 5th ed., John Wiley & Sons, 2001, pp. 427-511.

[5] W. S. Cleveland, S. J. Devlin, "Locally weighted regression: An approach to regression analysis by local fitting", Journal of the American Statistical Association, vol. 83, no. 403, pp. 596-610, Sep., 1988.

[6] D. M. Hamby, "A comparison of sensitivity analysis techniques", Health Physics, vol. 68, pp. 195-204, 1995.

[7] I. Kioutsioukis, S. Tarantola, A. Saltelli, D. Gatelli, "Uncertainty and global sensitivity analysis of road transport emission estimates" Atmos. Environ., vol. 38, pp. 6609–6020, 2004.

[8] R. Pastorelli, S Tarantola, M G Beghi, C. E. Bottani, A. Saltelli, "Design of surface Brillouin scattering experiments by sensitivity analysis", Surface Science, vol. 468, pp. 37–50, 2000.

[9] J-M Zaldivar, F. Campolongo, "An application of sensitivity analysis to fish population dynamics." In: *Saltelli et al., editors. Sensitivity analysis*, New York: Wiley, pp. 367–83, 2000.

[10] V. Czitrom, "One-Factor-at-a-Time Versus Designed Experiments". American Statistical Association, vol. 53, 1999.

[11] D. M. Hamby "A review of techniques for parameter sensitivity analysis of environmental models", Environmental Monitoring and Assessment. vol. 32, pp. 135-154, 1994.

[12] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, "Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index", Computer Physics Communications, vol. 181, 2010.

[13] P. A. Ekstrom, "Eikos- A Simulation Toolbox for Sensitivity Analysis", Degree project, Faculty of Science and Technology, Sweden, February 2005.

[14] C. Shannon, N. Petigara, S. Seshasai, ”A Mathematical Theory of Communication”, vol. 27, pp. 379–423, 1948.

[15] I. Kovacs., M. Topa, A. Buzo, M. Rafaila, G. Pelz, “Comparison of sensitivity analysis methods in high-dimensional verification spaces”, Acta Technica Napocensis Electronics and Telecommunications Journal, Technical University of Cluj-Napoca, vol. 57, no. 3, pp. 16-22, 2016.

[16] A. Iosub, J. Kirscher, M. Rafaila, A. Buzo, G. Pelz, L. Goras, “Simulation-based Approach to Application Fitness for an E-Bike”, IEEE SAS, 2016.

[17] A. Iosub, J. Kirscher, A. Buzo, G. Pelz, L. Goras, “On the influence of angle sensor nonidealities on the torque ripple in PMSM systems– an analytical approach”, IEEE SMACD, 2016.