MONITORING PRODUCTIVITY IN INDUSTRIAL WELDING PIPES BY MONTE CARLO METHOD

Miguel Luiz Ribeiro Ferreira, miguelluiz@lamis.uff.br
Diogo Constancio, dinit22@yahoo.com.br
LAMIS – Laboratório de Montagem Industrial e Soldagem. Departamento de Engenharia Mecânica. Escola de Engenharia.
Universidade Federal Fluminense – Rua Passo da Pátria, 156 – Sala – 265 – Bloco D – Ingá – Niterói – RJ – CEP: 24210-240.

José Luiz Ferreira Martins, jlfmtem@vm.uff.br
LAMIS – Laboratório de Montagem Industrial e Soldagem. Departamento de Engenharia Mecânica. Escola de Engenharia.
Universidade Federal Fluminense – Rua Passo da Pátria, 156 – Sala – 265 – Bloco D – Ingá – Niterói – RJ – CEP: 24210-240.

Abstract: *The aim of this article is to analyze the viability of the use of Monte Carlo Method to monitor the productivity in the MAG welding of industrial piping by using the data obtained from small samples and generating virtual data by simulation. The simulation method will be applied in two different sceneries: in the first scenery, known, the real model will be based on a representative sample of productivity indicators. However, in the second scenery, unknown, we will assume the function probability in order to represent the real model. The Monte Carlo Method will be used to generate the virtual data of the productivity indexes with the use of the softwares Bestfit and @ RISK. Bestfit will be applied so as to make the adjustment of the probability curves attributed to the group of input data, while the @ RISK will be responsible for the generation of the artificial random numbers from functions of probability initials.*

Key-words: *Welding. Monte Carlo Simulation. Productivity's Indicators.MAG.*

1. INTRODUCTION

Welding is a key manufacturing process in the piping construction industry and productivity in the industrial sector is closely related to the welding performance. Nonetheless, there is little diversity of productivity monitoring methods used in pipelines' welding. The most used methods take either the entire production or the daily labor involved throughout an enterprise into account. Therefore, these monitoring processes consume a lot of time and resources, which entails, in many cases, limitation, or even renunciation of their use by the construction company builders. Thus, the objective of this paper is to evaluate the applicability of the Monte Carlo method in monitoring the productivity of the weld of carbon steel pipes, in which the MAG process was being applied, by collecting small samples, from which data are generated by virtual simulation.

2. PRODUCTIVITY IN WELDING

According to Diekmann and Heinz (2001), productivity is classically defined by the Man hours per unit quantity produced. This relationship is commonly used in industry. As a consequence, the productivity in welding is generally defined as the amount of deposited weld metal for the amount of human resources consumed in this process. Still, productivity monitoring in welding is mentioned in literature by many indicators. There are indicators that consider only the open arc deposition and those that take into consideration the total execution time of the joint. The latter is the most used in the industry, as evidenced in the document metrics' industry standards PROMINP (2010), established with the participation of EPC industry in Brazil, in which welding indicators' standards are established.

In the case of the indicators that consider the total execution time of the joints, the majorities relate the weld volume, usually expressed in cm^3, or deposited mass, usually expressed in kilograms (kg), to the quantity of Man hours consumed in the welding operation.

Concerning the workforce, we can find the following conditions: amount of Man hours welders only; amount of welders and auxiliary personnel Man hours; amount of Man hours of welders, auxiliary personnel and welding supervision at the lowest level. In this work we use historical data consolidated on Gioia and Silva Junior (2007), which consider the productivity of each welder's identification mark in number of worked days, expressed in cm3/per Man hours. The workmanship considered in the calculation of indicators takes into account the actions linked to the welding activities exercised by the welder. On the other hand, it is considered that the observation time in which productivity is measured should begin with the welding joint and end with the final cleaning, subsequent to the finishing weld pass.

3. MONTE CARLO METHOD

In accordance with Shimizu apud Limmer (1997), the Monte Carlo method consists in generating artificial probability of occurrence of an event from a pre-established distribution lay using simulation. In this method the RANDOM function is responsible for generating random numbers, and the RANDOMIZE function is used to perform the simulations. The random numbers are generated from a stochastic experiment, in which the results are previously unknown or uncertain.

According to author, the random numbers correspond to values of a probability density function previously defined and expected to characterize through simulation. The probability density function shape is important and necessary to identify the random variables involved as discrete or continuous.

From MUN (2006), the Monte Carlo method is a powerful simulation parametric method, that is to say it is necessary to specify parameters for simulation before it starts, capable of providing solutions to difficult and complex practical problems in an easier way. Therefore, it represents an alternative to highly complex stochastic mathematical models. In its simplest form, the method is a random generator of pseudo-numbers or samples that can create millions of combinations of these values, resulting in multiple scenarios undertaken by the model, and allowing the analysis of their main characteristics. The values generated are used in the model in form of probability distributions that reflect the uncertainty associated to the variable. Nowadays, this method can be applied in analysis and quantification of risks as well as assisting the conduction of forecasts and estimative in multiple fields, such as research and development, and engineering.

The steps of the Monte Carlo method listed below were excerpted from the article *Aplicação do Método de Monte Carlo em Análise de Riscos em Projetos de Construção*, from Morano and Ferreira (2003).

I - Grouping Data

This step is used to group the interested data collected in a chart.

II - Frequency and Histogram

In this step, the data is grouped in a table composed by a number of classes, which are used as input data for the determination of frequency distribution and histogram construction. Therefore, two requirements are fundamental: the definition of the number of classes and consequently the interval value. In statistics there is no default unit set to establish the number of histogram classes. Although there are particular values, the authors admit the occurrence from 7 to 30 classes. Fonseca and Martins (1982) and Crespo (1996) argue that there is no formula for calculating the number of classes; however, both suggest the use of Sturges rule, which is shown in Eq. (1).

$$K = 1 + 3.22 \cdot \log N \tag{1}$$

Where: K = number of histogram classes; N = number of sample elements.

III - Generating Function Choice

The third step is the choice of the input distribution or generating function, according to which the input data will be distributed. As well as the last step, the input distribution selection in simulation is not a consensus among authors. Morano and Ferreira (2003) identified two distinct groups of authors. A group formed by names like Raftery (1994) and Kerzner (1998) argues that the choice of simulation distribution should be based on statistical analysis of the sample data. The second group, however, defends the choice of specific distributions, such as the Normal distribution, Beta, uniform, Triangular and Poisson. In Risk Management and Construction, from Flanagan and Norman (1993) various distributions are referred to as the most common. As a consequence, these functions should be easily identifiable, updatable when new data is added, and flexible to the shapes that they can assume. This step is considered the greatest difficulty in implementing an effective simulation. Data reliability, experience and professional knowledge are fundamental to the successful implementation of the methodology then.

IV – Definition of the Number of Simulations

The fourth step is the determination of the number of simulations, which varies from author to author. For instance, Bruni et al (1998) adopt from 200 to 1000 iterations, while Grey (1995) proposes the use of a minimum of 300 interactions. This quantities refer to the minimum necessary to represent the generating distribution considering that a greater number of simulations increases its reliability. Thus, this quantity should be chosen according to the situation and based on practical experience in implementing the method. In construction, 100 simulations are initially a reasonable number, yet they should be verified by chi-square testing in order to evaluate of the compliance degree between the input data distribution and the built distribution at the end of the simulation .

V - Result Analysis

The fifth and last stage of the simulation process is the analysis of results. The final step of the simulation is the development of a histogram and an accumulated frequency chart to analyze the resulting probability for shape distribution and cumulative frequency distribution. With the histogram ready, it is possible to distinguish important classes, such as those of higher occurrence of variables and those with the distribution curve defined, and also to calculate the probabilities of the occurrence of variables.

4. METHODOLOGY

4.1. Description

The experiment consists in evaluating the applicability of the Monte Carlo method in estimating productivity of welding pipelines from small data samples. Two scenarios were used in order to conduct the experiment: in the first one, the type generating function is known “a priori” (in advance), based on the analysis of the behavior of a real sample, whereas in the second the type of the generating function is not known. In this case, we used the Beta function, as recommended by literature. Three data samples were taken from a database with about 52 elements reporting the daily welders productivity, in (cm³/Hh). The procedure for setting these three data samples was conducted by selecting random numbers correlated to the data of the actual data available. The collection was done through manual drawing of the values, with reposition for n = 10, 15 and 20, respectively.

4.2. Determination of Generating Functions

Two simulation scenarios were taken into account in the determination of generating functions, in which only the tables of random numbers drawn above are known. The difference is that in the first one the kind of generating function is known from a sample containing the data of global productivity, obtained from a database by using the software ControlTub and considering an equivalent joint of 4 inches in diameter. Then, the selected data was applied to welders. This productivity measurement considers only the Man-hours of the welders and takes at least ten days of production into account. In the second scenario, however, it was assumed that this information does not exist, so Beta generating function was chosen, as recommended by several authors.

After we verified the behavior of the overall sample data containing the data collected in the field, it was shown that the curve that best fitted them was the Pearson 5 distribution. Therefore, this curve is considered the generating function for the first scenario mentioned above. The parameters characterizing the distribution are presented in the Tab. 1 below for the three established conditions.

Table 1 - Parameters of Pearson5 Distribution - samples with n = 10, 15 and 20 elements

Pearson5 Parameters			
Parameters / Statistics	**Table**		
	n=10	**n=15**	**n=20**
α	2,737331396	2,792079766	2,960613005
β	80,96659236	81,37655008	81,34191235
Mean (cm³/HH)	46,6	45,4	41,5
Std Dev (cm³/HH)	54,274	51,022	42,33

In the second scenario, we used the generating function in which the type of Beta distribution best represents the unknown sampling universe, as already mentioned. The parameters α1 and α2, which are shown in Tab. 2, were calculated in the same way as in the first scenario.

Table 2 – Parameters of Beta Distribution – Samples with n = 10, 15 and 20 elements

Beta Parameters – MAG			
Parameters / Statistics	Table		
	n=10	n=15	n=20
α1	0,656364803	0,710703216	0,879271093
α2	13,42750789	14,94044998	20,31411236
Mean (cm³/HH)	0,0466	0,0454	0,0415
Std Dev (cm³/HH)	0,0543	0,051	0,0423

The values of the mean and standard deviation were divided by 1000 because the Beta function admits domain between $0 \leq x \leq 1$. Without this procedure it would not be possible to determine the functions parameters.

4.3. Simulation Configuration

The configuration of the simulations in @ RISK was standardized, and it is valid for both scenarios. The basic parameters required to configure a simulation are: number of iterations, number of simulations, sampling data, seed random number generator and standard recalculation.

Table 3 – Simulation Configuration – Software @RISK 4.5

Simulation Settings	
Number of iterations	100, 1000 e 10000
Number of simulations	1
Sampling	Monte Carlo
Generating seed	Random
Standard Recalculation	Monte Carlo

The number of iterations corresponds to the virtual number or amount of generated virtual statistical calculations performed by the computer after generating a new number. The number of simulations indicates the desired number of repetitions of the set presentation generated, the seed indicates the value from which the numbers are generated, the type of sample data and the standard recalculation are based on the simulation method chosen, and related to the convergence results. The sampling type corresponds to the method by which the artificial values are generated in the Monte Carlo Method. The recalculation corresponds to the way the statistical calculations are updated, which happens to the extent that new artificial values are generated in the collections.

5. RESULTS ANALYSIS

5 .1 Simulation from Pearson5 Distribution

The artificial collections of productivity indicators presented below, in cm ³/Hh, were generated using the Pearson5 Distribution, respectively, for n = 10, 15 and 20, respectively. Tab 4 presents data for the sample with n = 10. Tab. 5 and Tab. 6 present sample data with n = 15 and n= 20, respectively.

Table 4 – Sample Statistics Generated by the Pearson5 Distribution for n = 10

Artificial Collection – MAG (Pearson5 Distribution real data and n=10)			
Statistics	**Iterations**		
	N=100	**N=1000**	**N=10000**
Min	10	7,1	5,2
Max	209	462,7	1376,8
Average	42,7	45	46,4
Mode	38,4	17	22,5
Median	34	33,7	33,5
Std Deviation	34,008	43,185	50,71

Table 5 – Sample Statistics Generated by the Pearson5 Distribution for n = 15

Artificial Collection – MAG (Pearson5 real data and n=15)			
Statistics	Iterations		
	N=100	N=1000	N=10000
Min	9,3	7,6	6,2
Max	194,4	432,5	1222,5
Average	45,9	44,3	46,1
Mode	21,1	19,1	20,8
Median	35,9	32,6	33,2
Std Deviation	32,598	42,034	51,423

Table 6 – Sample Statistics Generated by the Pearson5 Distribution for n = 20

Artificial Collection – MAG (Pearson5 Distribution real data and n=20)			
Statistics	Iterations		
	N=100	N=1000	N=10000
Min	9,5	8,4	6,1
Max	633,5	385,5	867,5
Average	50,4	41,7	41,5
Mode	44,8	22,5	11,1
Median	31,4	31,2	30,8
Std Deviation	78,212	38,593	39,429

5.2. Simulation from Beta Distribution

The artificial arrays of productivity indicators presented below, in cm³ per Man hours were generated using the Beta Distribution for n = 10, 15 and 20, respectively,. The statistical data are presented in Tabs. 7, 8 and 9.

Table 7 - Statistics of the Sample Generated by the Beta Distribution for n = 10

Artificial Collection – MAG (Beta Distribution and n=10)			
Statistics	Iteration		
	N=100	N=1000	N=10000
Min	0,00051	0,00065	0
Max	252,2	337,6	526,1
Average	40,3	46,8	46,8
Mode	0,02062	17,4	100,7
Median	24,9	28,6	27,6
Std Deviation	47,315	52,511	54,265

Table 8 - Statistics of the Sample Generated by the Beta Distribution for n = 15

Artificial Collection – MAG (Beta Distribution and n=15)			
Statistics	**Iteration**		
	N=100	**N=1000**	**N=10000**
Min	0,3552	0,0295	0,0168
Max	117,3	153,4	177,8
Average	25,4	26,4	26,4
Mode	22	8,4	10,2
Median	21,1	20,5	20,7
Std Deviation	21,695	21,903	21,477

Table 9 - Statistics of the Sample Generated by the Beta Distribution for n = 20

Artificial Collection – MAG (Beta Distribution and n=20)			
Statistics	**Iteration**		
	N=100	**N=1000**	**N=10000**
Min	1,3742	0,3088	0,0398
Max	80,8	146,5	159,3
Average	23,1	23	23,5
Mode	4,5	12,2	14,5
Median	20,1	18,3	18,6
Std Deviation	15,568	18,445	18,791

5.3. Cumulative Probability Density Function (F)

The analysis of results was performed by comparing the cumulative probability curves adjusted in *Bestfit*, which were associated to each numerical set artificially generated in the previous section, with the relative model curves from the main data sample containing 52 items. 10 intervals were selected to adjust the cumulative curves so as to standardize the test. The curves presented below were adjusted with the use of the chi-square test with assistance from the software *BestFit*.

The Figs. 1, 2, 3 represent the curves of the accumulated probability density function generated by simulation from the Pearson 5 distribution compared to that achieved from the real sample.

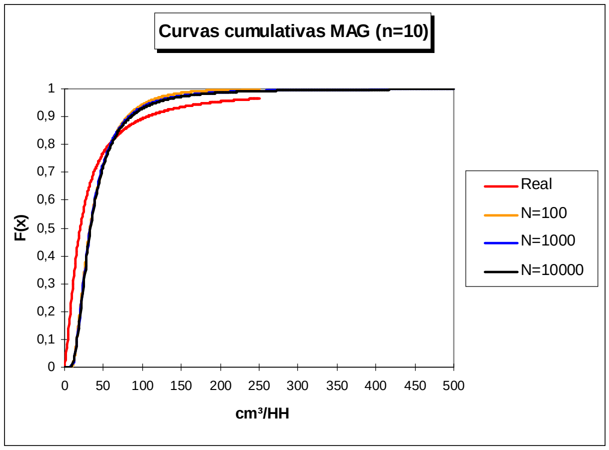

Figure 1 – Cumulative Probability Density Function (Pearson 5 Distribution with n = 10 and real curve)

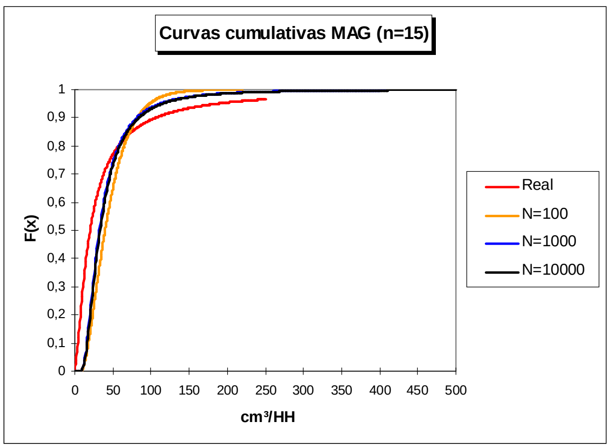

Figure 2 – Cumulative Probability Density Function (Pearson 5 Distribution with n = 15 and real curve)

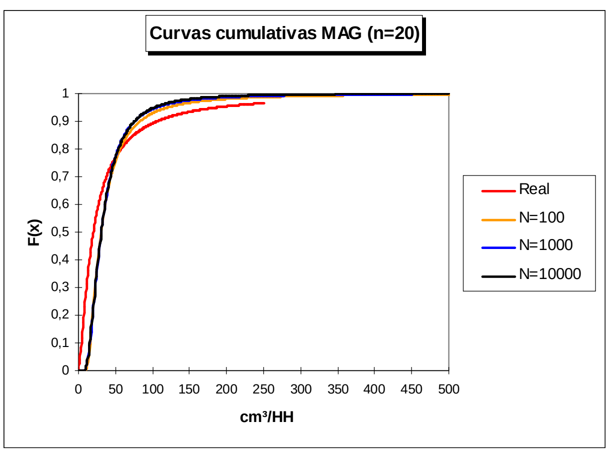

Figure 3 – Cumulative Probability Density Function (Pearson 5 Distribution with n = 20 and real curve)

Figures 4, 5 and 6 present the curves of the accumulated probability density function generated by simulation from the Beta Distribution, in a scenario in which the behavior of the population is not known, as recommended by the literature, compared to the real sample data.

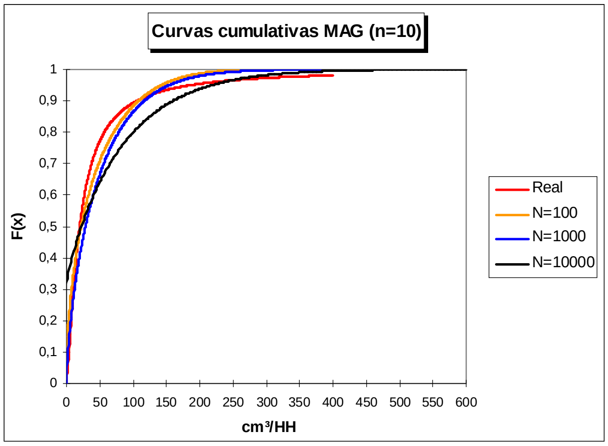

Figure 4 – Accumulated Probability Density Functions (Beta Distribution with n = 10 and real curve)

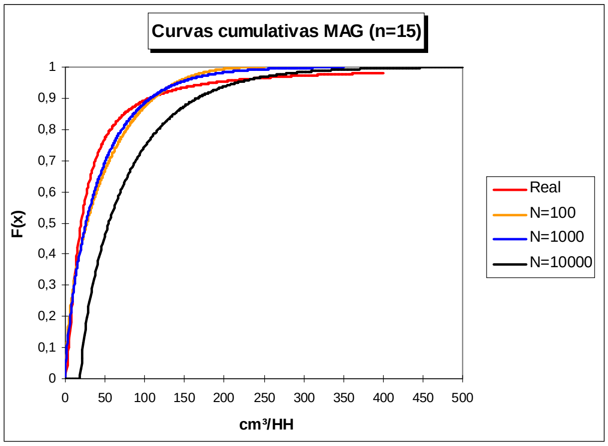

Figure 5 – Accumulated Probability Density Function (Beta Distribution with n = 15 and real curve)

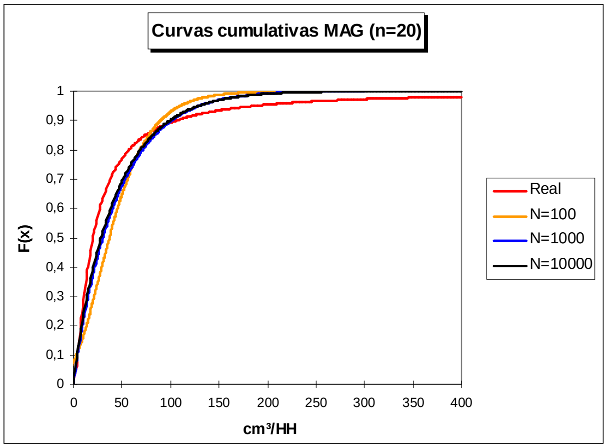

Figure 6 – Accumulated Probability Density Function (Beta Distribution with n = 20 and real curve)

The cumulative curves associated to the numerical sets generated by Beta Distributions, and determined according to the three tables of random numbers of the MAG process, presented a performance similar to those of the simulations using the generating functions of the Pearson5 type. Thus, it is possible to conclude that the use of Beta Distribution as the generating function yields similar results to those obtained with the generating Pearson5 Distribution, which best represents the sample universe data. Thus, the utilization of the Beta Distribution as the generating function to be used in the simulations, as recommended in literature, gives quite reasonable productivity estimates. This is an important fact, since in most of the real job situations we do not have preliminary information on the behavior of the productivity to be estimated. On the other hand, regarding the values adopted for estimating the productivity in concurrences or for job monitoring, this paper shows that the values near average presented a reasonable approximation in comparison with the sample universe data. However, the behavior of the data obtained by simulating the difference differs significantly from the sample universe when the maximum productivity achieved in the process is estimated. Therefore, it is necessary to estimate maximum productivity, which requires the adoption of values with a probability of occurrence located around the level 0.8 of the accumulated curve (F). Concerning the size of the sample (n) used to characterize the generating function of the simulation, the study reveals that there are reasonable results from $n = 15$ on.

6. CONCLUSION

Based on the analysis of the results we can conclude that, since a real model that represents the productivity behavior indicators with the dimension cm^3 per Man hours is not available yet, the probability Beta Distribution can be very useful in simulating generating functions as some authors affirm in the literature.

In the productivity estimation by Monte Carlo simulation, it is recommended to use a sample data with more than 15 elements in order to establish the generating function.

Furthermore, it seems to be a good practice to adopt values close to average to obtain considerably reasonable results. However, this does not occur for maximum productivity.

The data simulation obtained shows that it is possible to use Monte Carlo simulation to monitor productivity in welding based on small data samples, which allows the performance of welding monitoring during construction or even the implementation of corrective actions when necessary; although this practice is not common in business in building and construction industry nowadays.

7. BIBLIOGRAPHICAL REFERENCES

Bruni, A. L.; Famá, R., Siqueira, J. de O. “Análise do Risco na avaliação de projetos de investimento: Uma aplicação do Método de Monte Carlo”. Caderno de Pesquisas em Administração. São Paulo, v.1, n.6. 1998.

Crespo, A. A. “Estatística fácil”. 14. Ed. São Paulo: Saraiva, 1996. 224 p.

Diekmann, J.E, Heinz , J. “Determinants of Jobsite Productivity” CII – Construction Industry Institute, 2001. 144 p.

Flanagan, R., Norman, G. “Risk Management and Construction”. London: Blackwell Science, 1993. 208 p.

Fonseca, J. S. da; Martins, G. de A. “Curso de Estatística”**.** 3. ed. São Paulo: Atlas, 1987. 286 p.

Gioia, A.L.S., Junior, I. F. da Silva. “*Metodologia de medição de produtividade em soldagem de tubulações em obras industrias”*. Niterói, 2008. 36 p.

Grey, St. “Practical Risk Assessment for Project Management”**.** New York: John Wiley & Sons, 1995. 140 p.

Kerzner, H. “Project Management – A Systems Approach to Planning, Scheduling and Controlling”. Toronto, Canada: John Wiley & Sons, 1998. 1180 p.

Limmer, C.V. “Planejamento, orçamentação e controle de projetos e obras”. Rio de Janeiro: Livros e Técnicos e Científicos, 1997. 225 p.

Morano, C.A.R., Ferreira, M. L. R. “Aplicação do Método de Monte Carlo em Análise de Riscos em Projetos de Construção”. Niterói, 2003. 7 p.

Mun, J. “Modeling Risk - Applying Monte Carlo Simulation, Real Options Analysis, Forecasting and Optimization Techniques.”John Wiley & Sons, 2006. 591 p.

Prominp (Programa de Mobilização da Indústria do Petróleo e Gás Natural). Projeto E&P 27.5 – Métricas de Desempenho da Indústria Nacional. “Padrões de Métricas da Indústria EPC Nacional”. 2010.

Raftery, J. “Risk Analysis in Project Management”. London: E & FN SPON, 1994. 137 p.

8. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.