Uncertainty Analysis of Oil Well Flow Rate on the Basis of Differential Entropy

Ivan Luzyanin, Anton Petrochenkov and Sergey Bochkarev
Department of Microprocessor Units of Automation, Perm National Research Polytechnic University,
29 Komsomolsky ave. , Perm, Russia
{lis, pab}@msa.pstu.ru, bochkarev@msa.pstu.ru

Keywords: Uncertainty, Differential Entropy, Exponential Distribution Probability Density Function, Well Flow Rate, Oil Field, Oil Production.

Abstract: Oil well production efficiency depends on the accuracy of the flow rate prediction. The electrical submersible pumps are selecting and the well production control is carrying out based on the predicted values of flow rate. Inaccurate prediction may cause limitations of well deliverability or inefficient pumping. The prediction accuracy of flow rate changes in time related to initial data uncertainty that causes deviations between calculated flow rate values and measured ones. To minimize operating costs the same pump selection and control methods are used for groups of wells operating under the same conditions. However, sometimes wells demonstrate very different behavior even under the same conditions. In these wells flow rate changes becomes unpredictable by the common methods and additional studies required for correct prediction. The problem of finding wells with unpredictable flow rates at the early operation stages is very important because their inefficiency can significantly increase in time without special operation methods. The article considers the method of finding wells with potentially unpredictable flow rate changes with use of the entropy concept. The main feature of this method is that it is appropriate for data of any distribution types with given probability density function. The article discusses the relation between the value of joint reduction in uncertainty obtained from entropies of calculated flow rates and measured ones for a single well and the deviations between these flow rates. The novelty of the article is that the joint reduction in uncertainty in calculated value of well rate when knowing measured well rate is proposed as the measure of the well flow rate predictability.

1 INTRODUCTION

At present oil remains the main energy source for many fields of activity. However, its production becomes more difficult and energy intensive every year. This is caused both by oil reserves depletion and by increase in number of fields is being operated under complex geological conditions [1].

Today the main oil producing method is pumping with use of electric driven submersible pumps (ESP). ESP can work with high performance and efficiency in deep wells but its use in new oil fields and under complex geological conditions is limited. The limitations appears mainly when selecting the submersible equipment. The ESP energy efficiency is ultimately depends on accurate equipment selection for specific operating conditions. Incorrect selection can cause inefficient well operation for all pump life cycle (since the ESP replacement is carrying out only in case of its failure).

Large number of parameters are used when selecting ESP equipment. Some of them are found by statistical methods. The most significant parameter that determines energy efficiency is the desired well flow rate [2]. It is usually calculated under conditions of uncertainty and data incompleteness (especially in the early stages of field exploitation). Uncertainty is caused by the inability of detailed study of the reservoir and the data incompleteness is caused by experiment limitations. Various statistical models of wells are developed to solve this problem. Since each well has individual conditions, it requires an individual model. In practice, it is impossible to build individual models for each well, so the generalized models are used. The models are usually based on regression equations, their coefficients are found either with using experiments or by studying large datasets. These approaches are effective for well-studied fields that have been operating for a long time and for fields without special geological conditions. However, they do not consider geological factors that appear in individual wells and cause additional uncertainty of calculated parameters. This additional uncertainty together with data incompleteness can cause significant deviation between desirable flow rate and actual one. This in turn can be the reason of incorrect equipment selection or ineffective well production control. In addition, regression models usually require large amount of data that is unavailable at the early stage of well operation.

The article presents the investigation of the uncertainty that is existing in desirable (calculated with models) and actual (measured in well) flow rates of various oil wells with use of entropy concept. The ability of using information entropy for estimating the flow rates uncertainty for insufficiently known wells or for wells that operates in special regimes is studied. The main hypothesis is that when the entropies of both desirable and actual well flow datasets are known, the mutual information of these datasets will increase with decreasing flow rate predictability. This dependency will help to classify wells by flow rate predictability and select the most unpredictable wells for additional study.

Since the information entropy was originally introduced for discrete random variables [3], in this study the differential entropy of a continuous random variable is used instead. In general it is not an analogue of information entropy for continuous variables. However, when knowing the differential entropy, it is possible to obtain the mutual information for the case of continuous random variable.

When the above hypothesis is proven, the proposed method of predictability classifying will easily be applied in practice as it requires only knowing the data distribution law and allows data to have any distribution with given probability density function (PDF).

The research aim is to check the above hypothesis on the real data. The article considers the example of exponentially distributed data but the general algorithm of applying the method for any other distribution types is presented in the last part of the article.

The article includes four parts. First part presents short overview of commonly used flow rate calculation model and studies the causes of uncertainty. Second part considers data preparation and preliminary classification of statistical data obtained from oil fields. Third part considers determining the appropriate distribution law for classified datasets and presents the results of uncertainty analysis. The last part presents the generalized algorithm of applying the method for data of any distribution with given PDF.

2 CAUSES OF UNCERTAINTY IN WELL FLOW RATES

The ESP efficiency depends on current load of the motor that is represented by load factor (K_l). Its value can be calculated by (1).

$$K_l = \frac{N}{N_n} \quad (1)$$

where N is an electric power that is currently being consumed by ESP (found by (2)), N_n is rated ESP power. ESP efficiency has maximal value when K_l is one.

$$N = \frac{P_{pump} Q}{3600 \cdot 24 \cdot \eta_{pump} \cdot \eta_{motor}} \cdot 10^3 \quad (2)$$

In the above equation P_{pump} is a pump pressure required to lift oil to the surface, η_{pump}, η_{motor} are efficiencies of pump and motor respectively, Q is desirable (or actual) well flow rate. Pump pressure required for lifting oil to the surface and equipment efficiencies depend on well design and current operational regime. These parameters are usually constants for given regime. Therefore ESP load changes (and ESP efficiency) are ruled mostly by flow rate changes. These changes are typical not only for oil wells [4]. Moreover, desirable flow rate is used for ESP equipment selection. When pump has high performance and well has low flow rate the efficiency becomes significantly less than one. Nevertheless, in this case there is an ability to increase efficiency with using another regime. When pump has low performance and well has high flow rate, the efficiency will be low again but in this case efficiency increasing is more complicated than in previous one. Detailed description of these dependencies is given in [2].

Given considerations illustrate the significance of accurate calculation the desired well flow rate before well starts operating.

Standard well flow calculation model is based on Dupuit equation [2]. This equation represents the flow rate to the cylindrical well placed in the center of an “ideal” reservoir. “Ideal” reservoir must have regular geometry and be fully saturated with oil. Since there are no "ideal" reservoirs in real life, the equation is only useful for some sections in the real reservoir that fit the above requirements. These sections are usually separated from each other and have individual geometry. The production efficiency reaches its maximum if the reservoir can be divided into homogeneous sections of regular geometric shape with one or more wells operating in each section.

To find such sections, experimental data of similar fields are used. These data have uncertainty caused by experiment limitations and lack of information about field being studied. Analysis of the data obtained at the fields showed that the uncertainty of the reservoir structure and properties has a maximum value at the beginning of the field lifecycle and reaches the minimum at the end of its operation. Besides that, external factors such as rock destruction or changes in fluid properties also affect uncertainty [1], [5].

According to the Dupuit equation the deliverability of a given well is determined by productivity index (PI). In addition, flow rate depends on difference between reservoir pressure and bottomhole pressure (ΔP_f). PI and ΔP_f as well as reservoir geometry are either obtained experimentally or calculated with models.

Thus, the uncertainty of flow rate includes three components: the uncertainty of the reservoir geometry, the uncertainty of PI calculation and the uncertainty of ΔP_f calculation.

The actual flow rate value is measured by special sensors. The sensors have a measurement error that can usually be included in the rate value.

In these conditions, the comparative analysis of the uncertainties appearing in desirable and actual flow rates over long time periods can give significant results for understanding the ways of initial data uncertainty resolution.

It should be noted however that wells could have different operational conditions and work in different operational regimes. At that, desirable and actual flow rates must have different uncertainty.

3 PRELIMINARY DATA ANALYSIS

Statistical data for the study were obtained from 27 oil fields that are operating under different geological conditions. Obtained dataset includes 440 values of average annual well flow rates (220 values corresponds to desirable flow rates, others – actual ones).

At the first stage of the research the accuracy of predicting the actual flow rates was studied. For this purpose, the initial dataset was divided into subsets, each of that included the average annual values of the desirable and actual flow rates of a single well for all years of its operation. After that, the pairs of graphs (desirable rates changes in time and actual rates changes in time) were built for all wells. The graphs were classified according to the form of deviations of the desirable and actual flow rate curves. Figure 1 illustrates obtained classes of curves deviation.

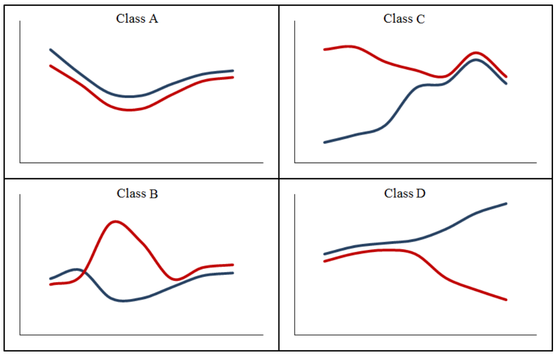

Figure 1: The classes of well flow curves deviations. Blue line corresponds to desirable rates, red line – actual rates.

The following classes were found:

- Class A – the desirable flow rate curve matches the actual flow rate curve;
- Class B – there is a single deviation inside desirable and actual flow rate curves;
- Class C – flow rate curves converges to the same shape;
- Class D – flow rate curves diverges from the same shape.

It should be noted that classification was built only by form of curves but not by value of deviations. For example class B includes both curves where desirable flow rates are below actual ones (as shown on figure) and vice versa. The classification tree (Figure 2) represents the probabilities of getting pairs of graphs into the classes A to D.

The tree includes two layers. The first one defines general form of discrepancy between graphs (classes A and B corresponds to generally coincident graphs; in opposite, graphs B and C correspond to not coincident ones). The second layer determines the belonging of the graphs to the specified class. The classification results can be interpreted as follows: the probability of accurate prediction of flow rate changes is 40% (class A), the probability of incorrect prediction of flow rate changes after some time period is 37% (classes B and D), the probability of incorrect prediction of flow rate changes in initial time period is 22 %. The overall probability of the prediction error is 59 %. Thus, the probability of accurate prediction is relatively small that possibly indicates the presence of large uncertainty in the initial data.

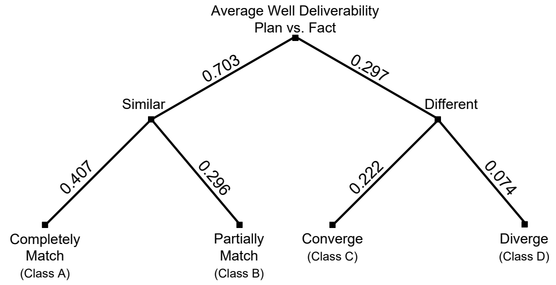

Figure 2: Classification tree for typical graph forms.

For further research, samples for each class and joint samples including samples of more than one class were obtained. Since the size of samples for classes C and D is small, they were combined in one sample. Table 1 includes samples that were used in the study for entropy analysis.

Table 1: Datasets used for uncertainty analysis (P – separate sets of desirable flow rates; F – separate sets of actual flow rates).

Classes included in sample (sample type)
A (P, F)
B (P, F)
C+D (P, F)
B+C+D (P, F)
A+B+C+D (P, F)

4 ENTROPY CALCULATION

The general (3) is commonly used for differential entropy calculation [3].

$$H(x) = -\int_S f(x)\log\left[f(x)\right]dx \quad (3)$$

where S is a support set of the random variable with given continuous distribution, $f(x)$ is a PDF for given X. The logarithm base defines the units of entropy. At the following study the base 2 is used, so the entropy is measuring in bits.

PDFs for statistical data were found by using the probability distribution histogram. The (4) was used for calculating the PDF value in each interval.

$$f_i(x) = \frac{m_i}{hN} \quad (4)$$

where m_i is a number of values from dataset that are included in the i-th interval, h is interval length, N is a number of values in sample. At the next step, the histogram was interpolated with PDF of appropriate theoretical distribution (Figure 3) and the distribution parameters were calculated. In the study it was hypothesized that all distributions have exponential distribution with PDFs given by (5) [6].

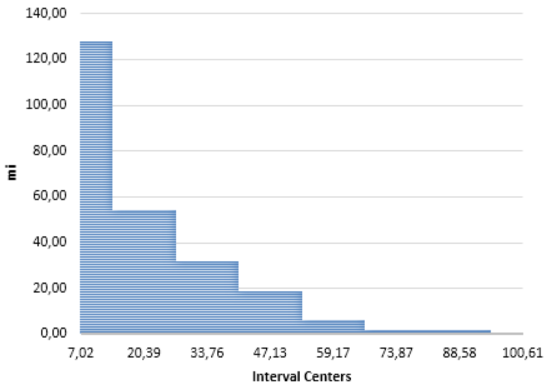

Figure 3: Histogram for probability density function and exponential probability density curve for dataset B+C+D FULL.

$$\begin{cases} f_X(x) = \lambda e^{-\lambda x} : x \geq 0 \\ 0 : x < 0 \end{cases}, \quad (5)$$

where λ is a distribution ratio obtained by the (6) [6]:

$$\lambda = \frac{1}{\overline{x}}, \quad (6)$$

where $\overline{x}$ is the mean value for a given dataset.

The hypotheses of the distribution were proven by F-test.

The distribution ratios for different samples are given in the Table 2.

Table 2: The values of the distribution ratio for different datasets.

Dataset	λ
A P	0.062
A F	0.055
B P	0.070
B F	0.071
C+D P	0.045
C+D F	0.051
B+C+D P	0.060
B+C+D F	0.060
A+B+C+D P	0.058
A+B+C+D F	0.057

Differential entropy for exponential distribution is obtained by the following (7) [7]:

$$H(X) = \log_2\left(\frac{e}{\lambda}\right), \quad (7)$$

The values of differential entropy for the datasets are given in Table 3.

Table 3: The values of the differential entropy for datasets.

Dataset	Differential entropy
A P	5.034
A F	4.945
B P	5.067
B F	5.014
C+D P	5.602
C+D F	5.401
B+C+D P	5.446
B+C+D F	5.351
A+B+C+D P	5.357
A+B+C+D F	5.464

The numeric value of differential entropy for a continuous distribution of random variable is not meaningful in practical tasks. Instead of this, the mutual information (I_{XY}) obtained from one random variable when given another random variable is the most important measure. For better understanding the interrelation between uncertainty and mutual information it is suggested to use term "joint reduction in uncertainty" in case of continuous variables [7]. This term is also more convenient for the following study because it describes the potential ability of reducing the uncertainty when obtaining datasets with different characteristics.

For two continuous random variables X and Y, the I_{XY} can be found by the following (8):

$$I_{XY} = -\int\int f_{XY}(x,y)\log_2\left(\frac{f_{XY}(x,y)}{f_X(x)f_Y(y)}\right)dxdy \quad (8)$$

where $f_{XY}(x,y)$ is a joint PDF of X and Y, $f_X(x)$, $f_X(y)$ are marginal distributions of X and Y respectively. If the differential entropies of the X and Y distributions are given, the (9) can be rewritten as follows:

$$I_{XY} = H_X + H_Y - H_{XY} \quad (9)$$

Here H_X and H_Y are the differential entropies of X and Y distributions themselves and H_{XY} is an entropy of joint distribution of X and Y that is obtained by the (10).

$$H_{XY} = -\int_X\int_Y f_{XY}(x,y)\log_2(f_{XY}(x,y))dxdy = \\ = -E(\log_2 f_{XY}(x,y)) \quad (10)$$

The above equation requires knowing the joint PDF of X and Y. For completely independent random variables the joint PDF is simply the product of PDFs for X and Y. If the variables are not independent, their dependency level is estimated by correlation coefficient ρ. In this case calculation of the joint PDF becomes more complicated as the dependency needs to be considered.

Since all samples in the table 1 have exponential distribution with PDFs given by (5), the joint PDFs or all the samples will be the PDFs of two exponentially distributed random variables and will have bivariate exponential distribution. Several studies consider the obtaining the bivariate exponential PDF for different cases [8-11]. In the following study the joint PDF is calculated as follows (11):

$$\begin{cases} f_{XY}(x,y) = \lambda_1\gamma_2 e^{-\lambda_1 x-\lambda_2 y-\lambda_3 y} : (x,y) \mid 0 < x < y \\ f_{XY}(x,y) = \lambda_2\gamma_1 e^{-\lambda_1 x-\lambda_2 y-\lambda_3 x} : (x,y) \mid 0 < y < x \\ f_{XX}(x,x) = \lambda_3 e^{-\lambda_3 x} : (x,y) \mid 0 < x = y \end{cases} \quad (11)$$

The (11) was obtained from general equation for bivariate exponential PDF given in [12] after some mathematical manipulations.

Parameters λ_1 and λ_2 in the equation are ratios of the corresponding PDFs for variables X and Y. λ_3 is the parameter that considers dependence between X and Y. It is obtained by the (12).

$$\rho = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3}, \quad (12)$$

where ρ is the correlation coefficient of X and Y.

Studying changes of I_{XY} for calculated and actual well flow rates in different samples allows finding the sets with the highest joint reduction of uncertainty (JR). These sets as hypothesized include wells with potentially the most unpredictable flow rates. In opposite, the sets with lowest JR values demonstrate the most predictable behavior. Therefore, it is expected that the JR value increases with decreasing of flow rate predictability. The predictability is estimated by deviations between calculated and measured well flow rates.

In the study the values of JR in uncertainty were calculated for datasets from the table 1. The calculation results along with the degree of dependency ρ are presented in Table 4.

The sample of full data including all classes from A to D has the relatively small JR value. This sample has a large part belonging to the classes A and B and a small part belonging to the other classes. In this case the uncertainty of the initial data is small and the data are high correlated. As a result there is not big uncertainty reduction for the calculated data when knowing the measured ones and this set has relatively good predictability. The JR value for data belonging to the class A itself is only a few smaller than this for class B. The maximal JR value was obtained for dataset of classes C and D. These classes include the potentially unpredictable wells.

Table 4: Joint reduction in uncertainty for pairs of datasets corresponding to calculated and actual (measured) values of flow rate.

Datasets	ρ	H_X+H_Y	H_{XY}	I_{XY}
A P + A F	0.952	9.979	4.781	5.198
B P +B F	0.914	10.081	4.619	5.462
C+D P + C+D F	0.576	11.003	0.0565	10.947
B+C+D P + B+C+D F	0.721	10.797	6.945	3.852
A+B+C+D P + A+B+C+D F	0.788	10.821	9.290	1.531

The described results are able to confirm the hypothesis of existing dependency between JR value and the predictability of well flow rate. It was found that datasets including data of wells that demonstrate deviations between calculated and actual values of flow rate had greater JR values then datasets without deviations. However, the amount of data used in the study is not able to give stable classification. In addition, the study shows that the algorithm is very sensitive to the concrete values of deviations between desirable and actual flow rates in any cases. This sensitivity is a cause of small difference between JR values of classes A and B. In the example presented in the article the concrete values of deviation are not considered.

It also should be noted that the preliminary classification of data was carried out by a single expert. So, the probability of misclassifying between classes A and B is relatively high. In practice it is suggested to use the algorithm presented in the next section for automatic classification based on the JR value.

4 APPLYING ENTROPY CONCEPT FOR ESTIMATION THE WELL FLOW PREDICTABILITY

The dependency between JR value and well flow rate predictability can be used for classification of wells by their flow rate predictability.

The initial dataset for this classification must include the equal amount of desirable (calculated with the models) and actual (measured in real well) flow rate values from any number of wells. The amount of values for each well must also be equal. Since oil production companies measure and recalculate the well rate values every equal time period (e.g. a year, a month etc.) these conditions are easy to match. The required total amount of data depends on how many classes need to be obtained. A number of classes (that determined by an expert before classification procedure) corresponds to a number of datasets are being obtained when classifying.

The classification procedure begins with dividing the initial dataset into parts of desirable and actual flow rate. After that, the distribution laws for each part are found and the parameters of distribution are calculated. Then the joint distribution law parameters are calculated. Finally, the entropies and the JR value for initial dataset are calculated. This value is used as a low constraint for JR value.

In the next iteration the initial dataset is divided into two equal subsets and the previous steps are repeated obtaining two JR values. The next iteration starts with comparing JR values. The dataset that has the lowest value is divided into two equal datasets and the previous steps are repeated again. The classification stops when the number of classes matches the value given by an expert.

As a result, the subsets sorted by the JR value are obtained. The subset with the largest JR value includes wells potentially the most unpredictable flow rates.

It should be noted that the division of the dataset in each iteration is carried out by the way that the values corresponding to one well cannot be divided into several subsets.

5 CONCLUSIONS

The concept of differential entropy presented in the article for estimating the predictability of oil well flow rates proves its usability. The dependency between JR value, calculated based on differential entropies of desirable and actual flow rates, and the flow rate predictability was obtained. The exponential distributed test dataset was used to illustrate work of the proposed method.

The described method of JR value calculation is appropriate not only for exponential distribution but also for the most of distribution types with given PDFs. However, it requires finding the joint PDF of two random variables that is sometimes a complicated task. Solutions of this task for different distributions are considered in [13]-[16].

Obtained dependency can be used for classification of the oil wells by their flow rates predictability. A simple iterative classification algorithm is presented in the article. The algorithm gives a solution of the problem of estimating the predictability of concrete wells at the early stages of their lifecycles. It will help oil production engineers and energetics to find wells that require special operation methods. In general, when using large amount of data describing flow rates of different wells the algorithm will help to correct the flow rate prediction models that are used for pump selection and well operation control.

The further study of the proposed algorithm will be carried out in field of estimation of the algorithm sensitivity and ways of its control. Additional study is also required for the procedures of finding the joint PDF for different distribution types.

The results of the study will be implemented in the software for analysis of the electrical power supply systems of oil fields [17].

The project is aimed at supporting of a new Master's program "Conceptual Design and Engineering to Improve Energy Efficiency" for preparing of engineers, scientific researchers and managers in energetics and related branches [18].

Research is also supported by educational and research grant 573879-EPP-1-2016-1-FR-EPPKA2-CBHE-JP by European program Erasmus+ (Project INSPIRE).

REFERENCES

[1] K. Bjørlykke, Petroleum geoscience: from sedimentary environments to rock physics, Springer-Verlag Berlin Heidelberg, 2010, 508 p.

[2] G. Takacs, Electrical submersible pump manual: design, operations, and maintenance, Gulf Professional Publishing, 2009, 420 p.

[3] T. M. Cover, J. A. Thomas, Elements of information theory, 2nd edition, Wiley, 2006, 748 p.

[4] V. N. Fashchilenko, , S. N. Reshetnyak, Resonant behavior of electric drives of mining machines, Gornyi Zhurnal, vol. 7, 2017, pp. 80-83, doi: 10.17580/gzh.2017.07.15.

[5] N. J. Hyne, Nontechnical guide to petroleum geology, exploration, drilling and production, 2nd edition, Pennwell Books, 2001, 575 p.

[6] N. L. Johnson S. Kotz, N. Balakrishnan, Continuous univariate distributions, 2nd edition, vol. 1, Wiley, 1994, 761 p.

[7] J. V. Mihalowicz et. al., handbook of differential ntropy, CRC Press, 2014, 220 p.

[8] I. Ghosh, A. Alzaatreh, A new class of bivariate and multivariate exponential distributions, Far east journal of theoretical statistics, vol. 50, issue 2, 2015, pp. 77-98, doi: 10.17654/FJTSMar2015_ 077_098.

[9] S. Nadarajah, D. Choi, Arnold and Strauss's bivariate exponential distribution – products and ratios, New Zealand journal of mathematics, vol. 35, 2006, pp. 189-199.

[10] S. K. Iyer, D. Manjunath, R. Manivasakan, Bivariate exponential distributions using linear structures, The Indian journal of statistics, vol. 64, series A, pt. 1, 2002, pp. 156-166

[11] D. Kundu, R. D. Gupta, Bivariate generalized exponential distribution, Journal of multivariate analysis, vol. 100, issue 4, 2009, pp. 581-593, DOI: 10.1016/j.jmva.2008.06.012

[12] B. M. Bemis, Some statistical inferences for the bivariate exponential distribution, dissertation, 1971, 116 p.

[13] A. Seijas–Macías, A. Oliveira, An approach to distribution of the product of two normal variables, Discussions mathematicae probability and statistics, vol. 32, 2012, pp. 87-99, DOI: 10.7151/dmps.1146

[14] Y. A. Tashkandy, M. A. Omair, A. Alzaid, Bivariate and bilateral gamma distributions, International journal of statistics and probability, vol. 7, no. 2, 2018, pp. 66-79.

[15] E. Furman, On a multivariate gamma distribution, Statistics and probability letters, vol. 78, 2008, pp. 2353-2360.

[16] S. Nadarajah, A. K. Gupta, Some bivariate gamma distributions, Applied mathematics letters, vol. 19, 2006, pp. 767-774

[17] I. Luzyanin , A. Petrochenkov, Practical Aspects of Software Developing for the System of Structural and Functional Analysis of Power Supply Systems in Oil Companies, Proceedings Of The 5th International Conference On Applied Innovations In IT, vol. 5, 2017, pp. 65-69, doi: 10.13142/KT10005.31, WOS: 000402660300 009

[18] A. Lyakhomskii et. al., Conceptual design and engineering strategies to increase energy efficiency at enterprises: Research, technologies and personnel, Proceedings of 2015 IV Forum Strategic Partnership of Universities and Enterprises of Hi-Tech Branches (Science. Education. Innovations), 2015, pp. 44–47, doi: 10.1109/IVForum.2015.7388249, wos: 000380529800015.