STATISTICAL MODELS FOR PREDICTING OF TELETRAPHIC PARAMETERS OF MARKOV CHAINS

Stanimir Sadinov, Ivelina Balabanova, Georgi Georgiev

Abstract: *The presented work is in the subject area "Teletraffic systems with Markov chains". The object of the study is the Markov chain of type M / M / c / k, where c = 25. A complete factorial experiment is accepted with managed object factors - Avg. Arrival Rate; Avg. Service Time and Max Station Capacity, and three levels of their variation. Modeling of the specified system was performed. The simulation was conducted under accepted numerical values of the factors aligned with parametric coding levels. Experimental data for the parameters of the object - Arrival Time and Exit System was obtained. An intellectual information analysis has been performed. Based on the experimental data, the content of the experiment plans is formed. Regression procedures have been applied to analyze and select the most appropriate plan to find statistical models for predicting the change in teletraffic parameters. Detailed results from the analysis of the predicted models are presented. The results include a set of coefficients of determination, surfaces of the response, lines of the same response, analysis of residues and others.*

Keywords: *Teletraffic System; Markov Chain; Plan of Experiment; Regression Analysis; Statistical Model; Teletraffic Parameters.*

ITHEA Keywords: *H. Information Systems, F. Theory of Computation.*

Introduction

There are a number of studies on the application of Markov chains as tools to build models and analyze data in various fields such as genetics, economics and others. As well as those exploring the significance of the order and the characteristics of higher order chains.

An example is research to find mathematical models of Markov chains with a big order retrieving knowledge of the human genome, studying at different depths of memory. In another study, a statistical analysis of high-order chains and a small number of parameters was considered, where statistical evaluations of parameters and tests for defined parametric hypotheses [2] were developed and analyzed. A similar study again explores the limited models with integer time series. These results contain probabilistic characteristics and statistic conclusions of two models - a chain of "s" order partial "r" connections and circuit with conditional order.

It is interesting to use Markov's chains as models for prediction of the parameters of the wind when generating energy by wind turbines. The models are analyzed on the basis of a comparison between the theoretical derivative characteristics and their empirically determined analogues. Here speed, direction and wind force are used to define the states of the circuit. The matrix of transitions is determined by estimating the maximum probability based on transient data. There are also studies performing statistical analysis at Markov's Discrete Chains, consisting of evaluation of a transition matrix, standard error, confidence intervals and others.

It is known to adapt the chains of Markov as a method Markov-Chain Monte Carlo. Its applications are related to the determination of the real distributions of the probabilities of parameters in a different volume of data analysis related to the estimation of the quantity and direction of distribution of environmentally harmful chemical agents, etc. The volume of studies focused on statistical study of telegraphic parametric status in Markov chains is limited. This article shows the possibility of estimating the approximate moments of entering and leaving packet data in information passage through different system states. The target object of the study is a Markov chain with a number of servers and a limited tail size M / M / 25 / k.

Simulation modeling of telegraphic system M / M / 25 / k

Markov chain M / M / c / k modeling was performed using a graphical user interface in the Java Modeling Tool (JMT) simulator environment. Number of server serving devices and system users, respectively, "c = 25" and "cust. = 1000" were selected. The study was carried out based on the attached types of experimental plans, respectively: Composite plan Hartley for3; B3 symmetric composition plan; D-optimal plan at m = 3; Symmetric quasi-D-optimal plan (Pesochinian plan) at m = 3 and unsymmetrical quasi-D-optimal plan at m = 3.

The system input parameters are defined as controllable object factors labeled as follows: x1 - Avg. Arrival Rate (lambda) [cust./s] - average arrival rate in the system; x2 - Avg. Service Time S [s] - average system attendance time and x3 - Max. Station Capacity k [cust.] - maximum queue capacity (waiting places in the system). Their numerical values have been adopted, aligned to the coding parameter levels shown in Table 1.

As a result of the simulation performed under specify groups of code combinations for each of the specified planes of the experiment, experimental data were obtained for the defined object output parameters: y1 - Arrival Time [s] - arrival time in the system and y2 - Exit System [s] - time for system exit.

Table 1. Levels of control factors of object

Variable levels	x_{1}, cust./s	x_{2}, s	x_{3}, cust.
-1	0.25	0.50	31
0	0.50	1.00	41
+1	0.75	1.50	51

Regression analysis procedures are applied to verify the adequacy of mathematical equations from zero (1), first (2) and second (3) degrees.

$$y = b_o + b_1x_1 + b_2x_2 + b_3x_3 \tag{1}$$

$$y = b_o + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 \tag{2}$$

$$y = b_o + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 \tag{3}$$

The procedures are performed on the advanced matrix experiments. Analyzing the values of the Ri2 coefficients in the regression results, the best plan of the experiment - B3 Symmetric Composition Plan was selected. A detailed study of the selected plan was carried out in the Statistica product.

Results of the regression analysis on the B_3 Symmetric Composition PlanB_3

In Fig. 1 is an expanded matrix of experiments analyzed type plan an experiment based on the highest-level regression model. Variable indices reflect the impact of one or the interaction between two or more controllable factors.

Data: SimetrichenB3 (12v by 14c)

	1 x1	2 x2	3 x3	4 x12	5 x13	6 x23	7 x123	8 x11	9 x22	10 x33	11 y1	12 y2
1	1	1	1	1	1	1	1	1	1	1	241,53	244,73
2	-1	1	1	-1	-1	1	-1	1	1	1	693,66	697,48
3	1	-1	1	-1	1	-1	-1	1	1	1	289,95	291,02
4	-1	-1	1	1	-1	-1	1	1	1	1	855,6	855,91
5	1	1	-1	1	-1	-1	-1	1	1	1	283,84	285,66
6	-1	1	-1	-1	1	-1	1	1	1	1	818,96	820,51
7	1	-1	-1	-1	-1	1	1	1	1	1	275,76	275,82
8	-1	-1	-1	1	1	1	-1	1	1	1	738,2	738,69
9	1	0	0	0	0	0	0	1	0	0	270,5	270,53
10	-1	0	0	0	0	0	0	1	0	0	782,95	785,21
11	0	1	0	0	0	0	0	0	1	0	417,25	418,15
12	0	-1	0	0	0	0	0	0	1	0	459,67	459,77
13	0	0	1	0	0	0	0	0	0	1	407,96	407,98
14	0	0	-1	0	0	0	0	0	0	1	440,95	442,09

Fig.1. Expanded Experiment Matrix

The results of the nalysis are shown in Figures 2 through 4. In terms of the Arrival Time parameter for the zero, first and second degree models, the coefficients of determination are R^2 = 0.93949388, R^2 = 0.96258631 and R^2 = 0.99953506 (Fig.2.a); Fig.3.a) и Fig4.a). With the Exit System parameter, the criteria are equal to R^2 = 0.93921893, R^2 = 0.96181636 and R^2 = 0.99942524 (Fig.2.b); Fig.3.b) and Fig.4.b).

	Regression Summary for Dependent Variable: y1 (SimetrichenB3) R= ,96927492 R?= ,93949388 Adjusted R?= ,92134204 F(3,10)=51,758 p<,00000 Std.Error of estimate: 64,308					
N=14	b*	Std.Err. of b*	b	Std.Err. of b	t(10)	p-value
Intercept			498,341	17,18708	28,9951	0,000000
x1	-0,966884	0,077786	-252,779	20,33603	-12,4301	0,000000
x2	-0,062707	0,077786	-16,394	20,33603	-0,8062	0,438904
x3	-0,026396	0,077786	-6,901	20,33603	-0,3393	0,741366

a)

	Regression Summary for Dependent Variable: y2 (SimetrichenB3) R= ,96913308 R?= ,93921893 Adjusted R?= ,92098461 F(3,10)=51,508 p<,00000 Std.Error of estimate: 64,504					
N=14	b*	Std.Err. of b*	b	Std.Err. of b	t(10)	p-value
Intercept			499,539	17,23931	28,9768	0,000000
x1	-0,967003	0,077962	-253,004	20,39783	-12,4035	0,000000
x2	-0,059120	0,077962	-15,468	20,39783	-0,7583	0,465758
x3	-0,025092	0,077962	-6,565	20,39783	-0,3218	0,754195

b)

Fig.2. Results using model (1) for a) parameter y_1 and b) parameter y_2

N=14	Regression Summary for Dependent Variable: y1 (SimetrichenB3) R= ,98111483 R?= ,96258631 Adjusted R?= ,91893700 F(7,6)=22,053 p<,00070 Std.Error of estimate: 65,284					
	b*	Std.Err. of b*	b	Std.Err. of b	t(6)	p-value
Intercept			498,341	17,44786	28,5617	0,000000
x1	-0,966884	0,078966	-252,779	20,64459	-12,2443	0,000018
x2	-0,062707	0,078966	-16,394	20,64459	-0,7941	0,457382
x3	-0,026396	0,078966	-6,901	20,64459	-0,3343	0,749546
x12	0,017465	0,078966	5,105	23,08135	0,2212	0,832292
x13	-0,008647	0,078966	-2,527	23,08135	-0,1095	0,916373
x23	-0,127953	0,078966	-37,400	23,08135	-1,6204	0,156283
x123	0,079628	0,078966	23,275	23,08135	1,0084	0,352189

a)

N=14	Regression Summary for Dependent Variable: y2 (SimetrichenB3) R= ,98072237 R?= ,96181636 Adjusted R?= ,91726878 F(7,6)=21,591 p<,00075 Std.Error of estimate: 66,003					
	b*	Std.Err. of b*	b	Std.Err. of b	t(6)	p-value
Intercept			499,539	17,64001	28,3185	0,000000
x1	-0,967003	0,079774	-253,004	20,87194	-12,1217	0,000019
x2	-0,059120	0,079774	-15,468	20,87194	-0,7411	0,486619
x3	-0,025092	0,079774	-6,565	20,87194	-0,3145	0,763763
x12	0,017161	0,079774	5,020	23,33554	0,2151	0,836798
x13	-0,008512	0,079774	-2,490	23,33554	-0,1067	0,918502
x23	-0,126650	0,079774	-37,047	23,33554	-1,5876	0,163475
x123	0,078679	0,079774	23,015	23,33554	0,9863	0,362090

b)

Fig.3 Results using model (2) for a) parameter y_1 and b) parameter y_2

	Regression Summary for Dependent Variable: y1 (SimetrichenB3) R= ,99976750 R?= ,99953506 Adjusted R?= ,99798525 F(10,3)=644,94 p<,00009 Std.Error of estimate: 10,292					
N=14	b*	Std.Err. of b*	b	Std.Err. of b	t(3)	p-value
Intercept			432,476	6,559974	65,9265	0,000008
x1	-0,966884	0,012449	-252,779	3,254658	-77,6668	0,000005
x2	-0,062707	0,012449	-16,394	3,254658	-5,0371	0,015084
x3	-0,026396	0,012449	-6,901	3,254658	-2,1203	0,124142
x12	0,017465	0,012449	5,105	3,638819	1,4029	0,255221
x13	-0,008647	0,012449	-2,527	3,638819	-0,6946	0,537270
x23	-0,127953	0,012449	-37,400	3,638819	-10,2781	0,001964
x123	0,079628	0,012449	23,275	3,638819	6,3963	0,007740
x11	0,192697	0,013412	94,249	6,559974	14,3672	0,000731
x22	0,012234	0,013412	5,984	6,559974	0,9122	0,428940
x33	-0,016400	0,013412	-8,021	6,559974	-1,2228	0,308718

a)

	Regression Summary for Dependent Variable: y2 (SimetrichenB3) R= ,99971258 R?= ,99942524 Adjusted R?= ,99750939 F(10,3)=521,66 p<,00012 Std.Error of estimate: 11,452					
N=14	b*	Std.Err. of b*	b	Std.Err. of b	t(3)	p-value
Intercept			432,819	7,299246	59,2964	0,000011
x1	-0,967003	0,013841	-253,004	3,621440	-69,8628	0,000006
x2	-0,059120	0,013841	-15,468	3,621440	-4,2712	0,023557
x3	-0,025092	0,013841	-6,565	3,621440	-1,8128	0,167519
x12	0,017161	0,013841	5,020	4,048893	1,2398	0,303175
x13	-0,008512	0,013841	-2,490	4,048893	-0,6150	0,582068
x23	-0,126650	0,013841	-37,047	4,048893	-9,1500	0,002760
x123	0,078679	0,013841	23,015	4,048893	5,6843	0,010791
x11	0,194189	0,014912	95,051	7,299246	13,0221	0,000978
x22	0,012547	0,014912	6,141	7,299246	0,8414	0,461924
x33	-0,015902	0,014912	-7,784	7,299246	-1,0664	0,364456

b)

Fig.4. Results using model (3) for a) parameter y_1 and b) parameter y_2

In the first two models the significant coefficients of regression b_i alphabetically coincide for both output parameters, respectively b_0 and b_1. Whereas in the regression model (3) the coefficients of significance are $b_0 = 432.476$, $b_1 = -252.779$, $b_2 = -16.394$, $b_{23} = -37.400$, $b_{123} = 23.275$ and $b_{11} = 94.249$ for y_1; $b_0 = 432.819$, $b_1 = -253.004$, $b_2 = -15.468$, $b_{23} = -37.047$, $b_{123} = 23.015$ и $b_{11} = 95.051$ with y_2. Very good indicators of Fisher's criteria have been achieved $F(10;3) = 644.94$ and $F(10;3) = 521.66$ and their respective probabilities $p < 0.00009 << 0.05$ and $p < 0.00012 << 0.05$. Where 0.05 is the assumed level of coefficient α for the parameters y_1 and y_2.

The numerical values of the coefficients of determination in equation (3) for both output parameters approximate to the maximum theoretically the ideal best value of this coefficient "1". That is why (3) is chosen as the basis for defining the final statistical models describing the most appropriate behavior of the object of study. System predictive models based on applied regression analyzes can be recorded as follows:

$$y_1 = 432.476 – 252.779x_1 – 16.394x_2 - 37.400x_2x_3 + 23.275x_1x_2x_3 + 94.249x_1^2 \quad (4)$$

$$y_2 = 432.819 – 253.004x_1 – 15.468x_2 - 37.047x_2x_3 + 23.015x_1x_2x_3 + 95.051x_1^2 \quad (5)$$

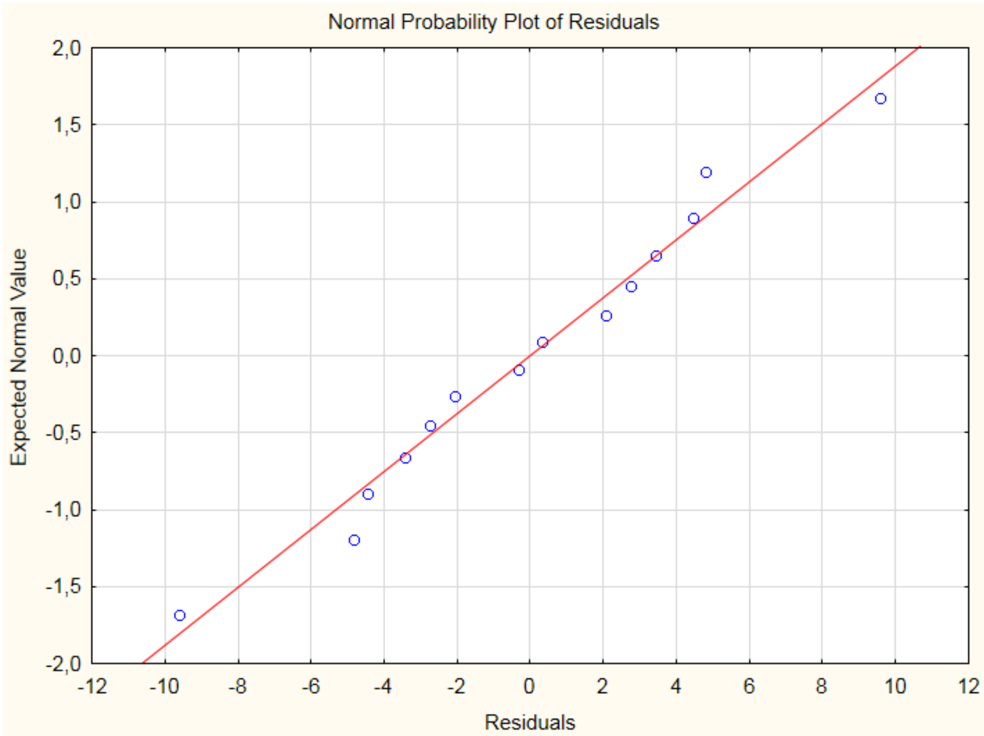
Analysis of the residues for the selected for systematic forecasting models

The "normal probability charts of residues" presented in Figure 5 were generated. There is a very good residual approximation to the right angle of 45 for both output parameters. Residues have characteristics of normally distributed data supporting the correctness of the application of regression analysis when investigating the object telegraph system.

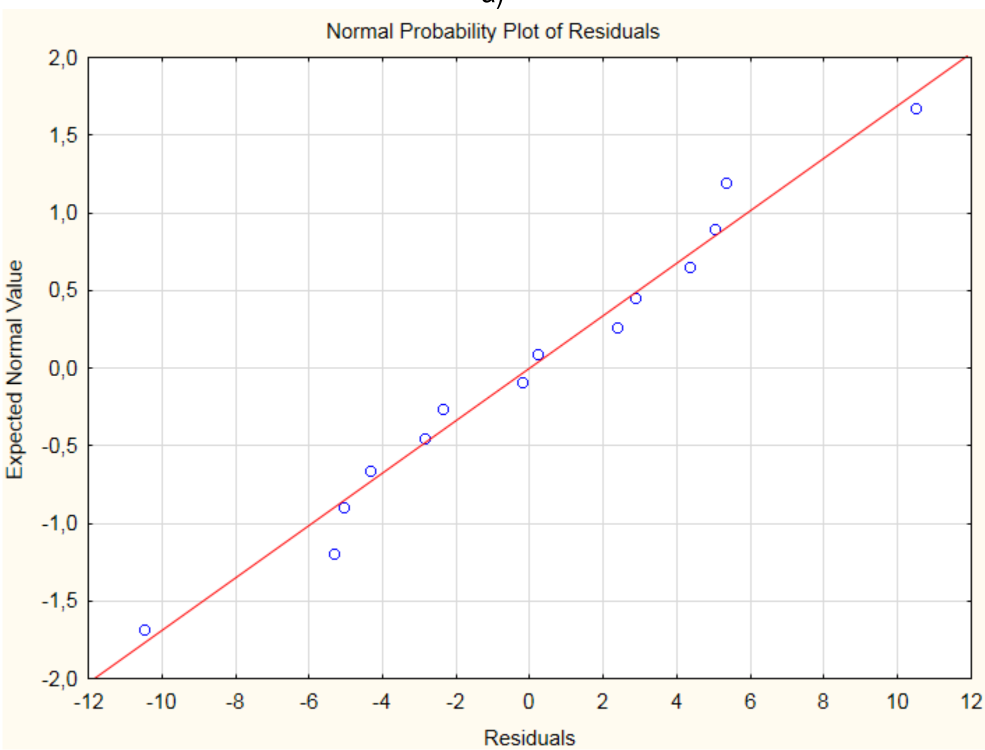
Surface response and lane response study in Equation (3) on *B_3* Symmetric Composition Plan

A three-dimensional graphical space is defined and "response surfaces" are constructed and their "equal response" surface sections are aligned with planes parallel to the x1Ox2, x1Ox3 and x2Ox3 planes. Graphical dependencies illustrate the areas of change in the controllable factors in which its output parameters are at the highest levels. Below are the "response replies" (fig.6) and "equal response lines" (fig. 7) regarding the x2Ox3 plane for the Arrival Time and Exit System parameters.

By analyzing the "equal response lines" in term of x1Ox2 and x1Ox3 planes for the parameters y1 and y2, they can be said to have the highest values at all the variation levels of the factors x2 and x3 and the low change levels of factor x1. In x2Ox3 plane, things look quite differently, there is a trend of large parameter changes only at extreme high for x2 and high variance levels for factor x3.

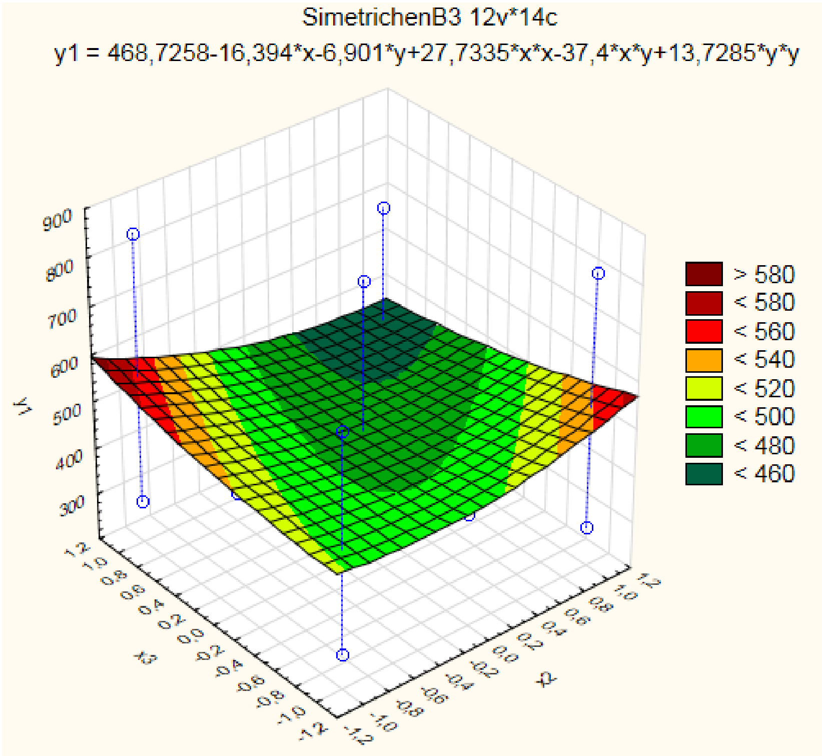


a)

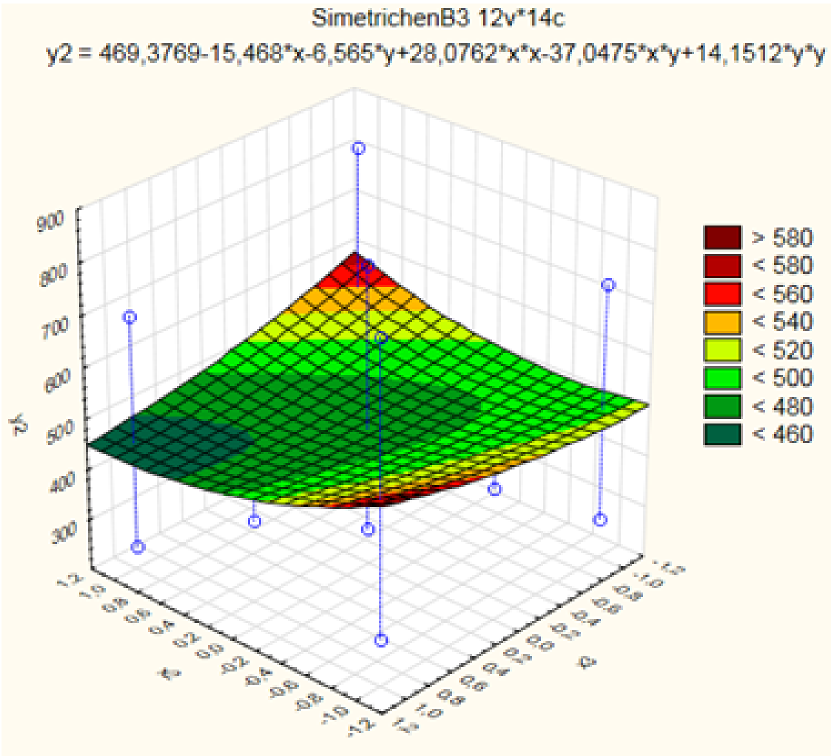


b)

Fig.5. Normal probability graphs when applying a) model (4) for parameter y_1 and (b) a model (5) for parameter y_2, concerning a B_3 Symmetric Composition Plan B_3

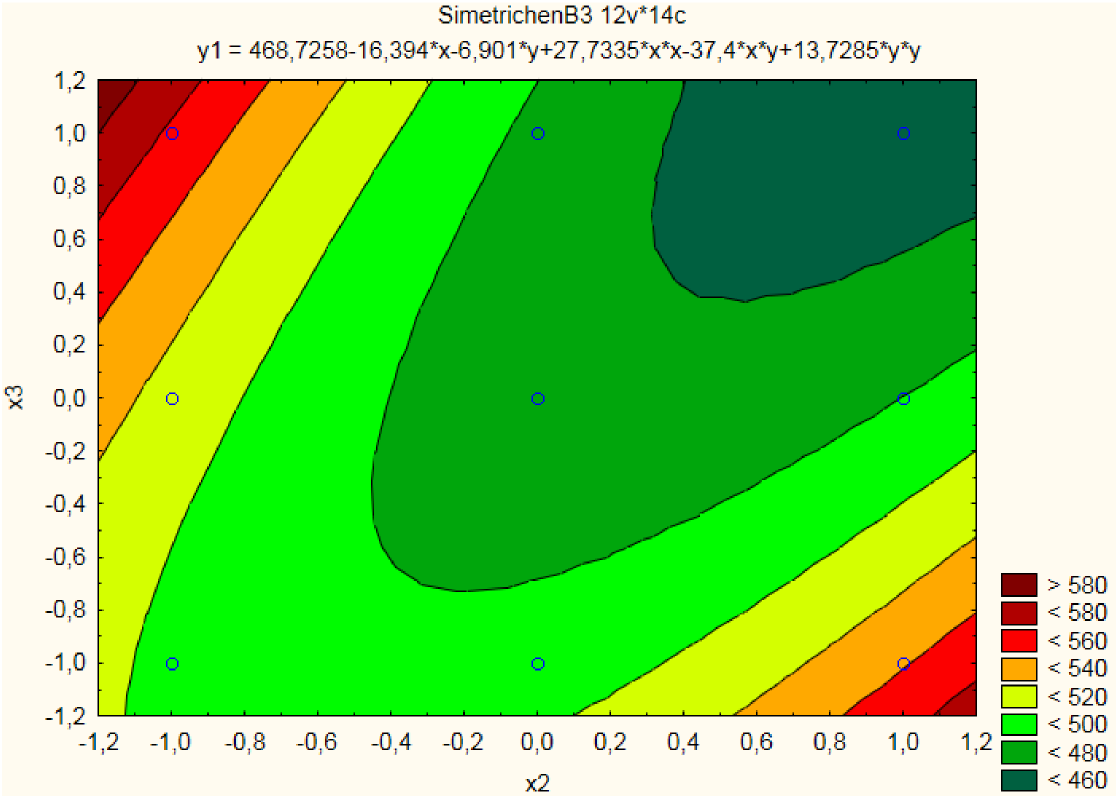


a)

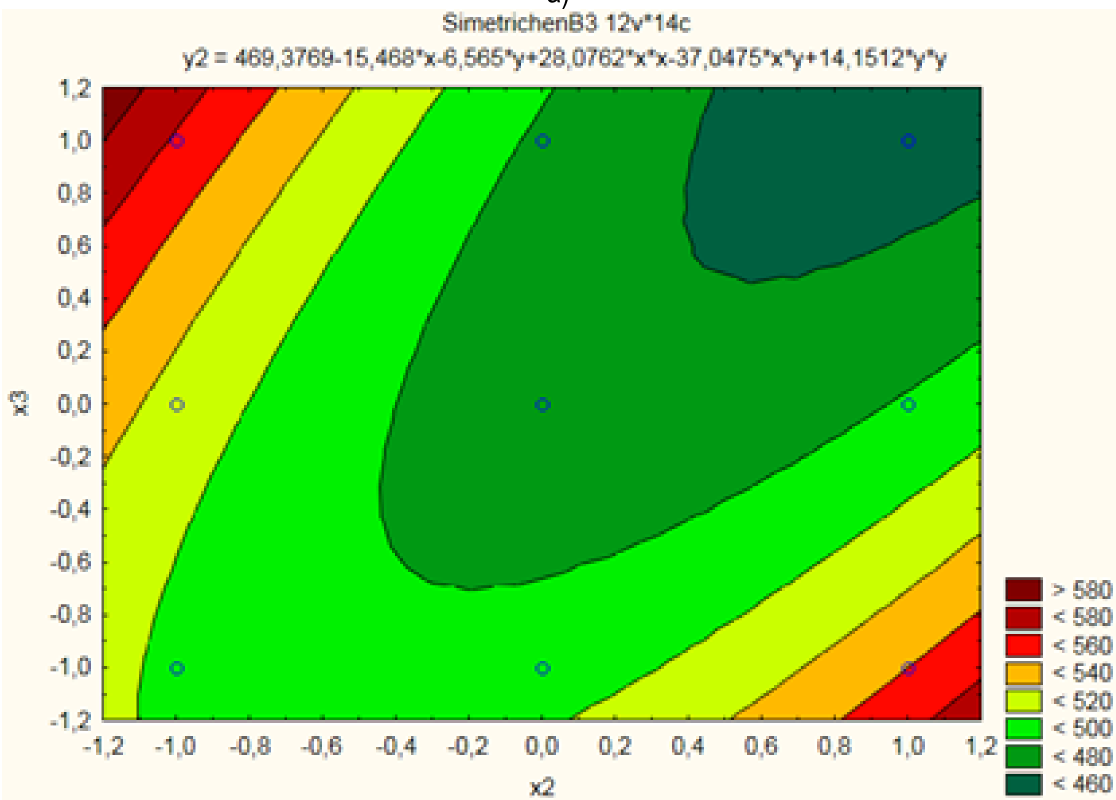


b)

Fig.6. Parameter Response Surfaces a) $y_1 = f(x_2,x_3)$ and b) $y_2 = f(x_2,x_3)$



a)



b)

Fig.7. Lines of the same response for parameters - a) $y_1 = f(x_2,x_3)$ and b) $y_2 = f(x_2,x_3)$

Conclusion

The tools of regression analysis have been successfully used in finding statistical models for mathematical prediction of the state of targeted teletraffic parameters. This gives prerequisites for future perspectives of the development of the presented paper in the field of communications and information systems. It is envisaged to study the specifics of the other types of Markov chains, and also studying the feasibility of regression equations of third or higher degree in the formation of Teletraffic predictive modeling.

Bibliography

[Kharin and Maltsew, 2010]] Kharin Yu., Maltsew M., On statistical analysis of Markov Chains with conditional memory depth. Electronic library of BSU, EBU BSU: NATURAL AND EXACT SCIENCES: Mathematics, EBU BSU: INTER-BRANCH PROBLEMS: Statistics, Vol.2. Publishing center of BSU, 2010, pp.22-25, ISSN: 978-985-476-848-9 (on-line), http://elib.bsu.by/handle/123456789/23927

[Kharin and Maltsau, 2014] Kharin Yu., Maltsau M., Markov chain of conditional order: Properties and Statistical Analysis. Austrian Journal of Statistics, Vol. 43/3-4, 2014. pp. 205-216. http://www.ajs.or.at/

[Kharin and Maltsew, 2017] Kharin Yu., Maltsew M., Statistical analysis of high-order Markov dependemcies. ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA, Vol.1, Number 1, 2017. pp. 79-91. http://acutm.math.ut.ee

[Lopes et al, 2012] Lopes V., Scholz T., Estanqueiro A., Novais A., On the use of Markov chain models for the analysis of wind power time-series, *11th International Conference on Environment and Electrical Engineering*. Environment and Electrical Engineering (EEEIC), Venice, Austria, 2012, pp. 770-775. IEEE, DOI 10.1109/EEEIC.2012.6221479 http://ieeexplore.ieee.org/document/6221479/

[Spedicato, 2017] Spedicato G., Discrete time Markov chains with R. The R Journal, Vol.9/2. CONTRIBUTED RESEARCH ARTICLE, 2017. pp. 84-104. ISSN: 2073-4859, https://journal.r-project.org/archive/2017/RJ-2017-036/

[Görlitz et al, 2011] Görlitz L., Gao Zh., Schmitt W., Statistical analysis of chemical Transformation kinetics using Markov-chain Monte Carlo methods. *Environ. Sci. Technol.*,, Vol. 45(10). Environmental Science and Technology, American Chemical Society, 2011. pp. 4429-4437. ISSN: 0123-0123 (print version), doi: 10.1021/es104218h https://pubs.acs.org/doi/abs/10.1021/es104218h

[Raja et al, 2017] , Raja A. , Mikael B., Jorge M., Sayyed M., Joel S., Syed Z., Raja A., Lars A., VMCMC: A graphical and statistical analysis tool for Markov chain Monte Carlo traces. BMC Bioinformatics, 2017. pp. 2-8. Open Access, DOI: 10.1186/s12859-017-1505-3 https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-1505-3

Authors' Information

Stanimir Sadinov – *Head of Department of the Communication Equipment and Technologies, Technical University of Gabrovo, Bulgaria; e-mail: sadinov.tc@abv.bg*

Major Fields of Scientific Research: Communication Networks and Systems, Digital Television Broadcasting, Mobile Communications

Ivelina Balabanova – *Department of the Communication Equipment and Technologies, Technical University of Gabrovo, Bulgaria; e-mail: ivstoeva@abv.bg*

Major Fields of Scientific Research: Communication Networks and Systems, Optical Communications, Teletraffic Systems, Automation of Design and Modeling

Georgi Georgiev – *PhD student in Technical University of Gabrovo, Bulgaria; e-mail: givanow@abv.bg*

Major Fields of Scientific Research: Measurement and control of air environment parameters in closed premises.