Annales Societatis Geologorum Poloniae (2010), vol. 80: 163–166.

TESTING THE RANDOMNESS OF LITHOSTRATIGRAPHIC SUCCESSIONS WITH THE MARKOV CHAIN METHODS

Marek DOKTOR, Andrzej KRAWCZYK & Wojciech MASTEJ

AGH University of Science and Technology, Faculty of Geology, Geophysics and Environmental Protection, al. Mickiewicza 30, 30-059 Kraków, Poland; e-mails: doktor@agh.edu.pl; akraw@geolog.geol.agh.edu.pl; wmastej@agh.edu.pl

Doktor, M., Krawczyk, A. & Mastej, W., 2010. Testing the randomness of lithostratigraphic successions with the Markov Chain methods. *Annales Societatis Geologorum Poloniae*, 80: 163–166.

Abstract: The crucial role in the studies on layer successions in lithostratigraphic sections with the application of stochastic model of depositional processes represented by the Markov chain is played by correct estimation of the matrix of transition numbers between lithofacies expected in a random sequence. Methods known from the literature are iteration procedures, which do not ensure quick, general solution. Hence, we propose a universal method based upon the Monte Carlo simulation technique. This method enables the researcher to estimate precisely and reliably the expected matrix of facies transitions.

Key words: Markov chain methods**,** Monte Carlo methods, coal-bearing successions.

Manuscript received 17 April 2010, accepted 25 June 2010

INTRODUCTION

Studies on layer successions in lithostratigraphic sections with the stochastic model of depositional process represented as discrete Markov process (Markov chain) have quite a long history. The early publications appeared in the 1960s (see Vistelius & Faas, 1965; Vistelius & Feygelson, 1965; Carr *et al.*, 1966; Krumbein, 1967; Potter & Blakely, 1967, 1968). The following years were devoted to intensive methodological studies, which enabled the researchers to recognize the specific character of geological applications of the Markov chains and to develop methods of correct interpretation of calculated results. Recently, the Markov chains are regarded as almost a standard tool of sedimentological analysis but still many publications appear, whose authors do not fully understand the applied mathematical methods and, thus, their conclusions are not fully correct. Therefore, it seems reasonable to present once again the principles of Markov model applications to the analysis of lithostratigraphic successions.

In order to test the lithostratigraphic section with the Markov chains, it must be transformed into a sequence $\{W_k\}$, $k = 1, 2, \ldots, N - 1$, of observed lithological varietes (lithofacies) $L_1, L_2, \ldots, L_M$. Two methods of transformation are used: recording the succeeding layers or recording the lithofacies distinguished at equal distances within the studied section. However, each metod has some limitations: in the former method information is lost on thicknesses of the recorded layers, in the latter one ambigous information appears due to the lack of precise criteria for selection of distances between the recorded lithofacies. The latter problem seems to be more serious (probably because the thickness analysis can be easily run with other methods); hence, the researches prefer rather the direct recording of the sequences of lithological varieties.

The main goal of the analysis of layer successions in lithostratigraphic sections is the recognition whether depositional process recorded in the section is random or whether any regularities are observed, expressed by preferred occurrence of some successions of lithofacies. Direct recognition of such preferences is quite easy, as computer techniques enable us to search quickly even very long lithostratigraphic sections. Unfortunately, the results of such studies cannot be sensibly interpreted due to the lack of relevant statistical methods.

The useful alternatives are methods resulting from the Markov chains theory. Although these methods cannot answer all questions interesting to a geologist, there exists an adequate theoretical basis which enables one to solve the most important problem, *i.e.*, the randomness of a given succession, as the rejection of randomness hypothesis justifies the undertaking of further, more specialized sedimentological studies.

The succession of elements in sequence $\{W_k\}$, which represents the succession of layers in the analysed lithostratigraphic section can be expressed as a square matrix:

$$\boldsymbol{F} \quad \begin{matrix} f_{11} & f_{12} & f_{1j} & f_{1m} \\ f_{21} & f_{22} & f_{2j} & f_{2m} \\ & & & \\ f_{i1} & f_{i2} & f_{ij} & f_{im} \\ & & & \\ f_{m1} & f_{m2} & f_{mj} & f_{mm} \end{matrix} \qquad (1)$$

where each element f_{ij} is equal to the number of events in the sequence $\{W_k\}$, and W_k = $\mathrm{L_i}$ and W_{k+1} = $\mathrm{L_j}$; more precisely, it is equal to the number of transitions of lithofacies $\mathrm{L_i}$ into lithofacies $\mathrm{L_j}$. In general, all values f_{ij} can be higher than 0 but, if during the recording of rocks in a given section only the succeeding lithological varieties are recorded, the transitions $\mathrm{L_i} \rightarrow \mathrm{L_i}$ (*i.e.*, between two following, lithologically identical layers) will become forbidden by definition and the tally matrix will evolve into:

$$\boldsymbol{F} \quad \begin{matrix} 0 & f_{12} & f_{1j} & f_{1m} \\ f_{21} & 0 & f_{2j} & f_{2m} \\ & & & \\ f_{i1} & f_{i2} & 0 & f_{im} \\ & & & \\ f_{m1} & f_{m2} & f_{mj} & 0 \end{matrix} \qquad (2)$$

It must be emphasized that this case is essentially different from that, in which 0 values appear in the matrix $\boldsymbol{F}$ as an empirical fact resulting from the lack of relevant transitions in the studied succession.

If a lithostratigraphic section recorded in the sequence $\{W_k\}$ is of random character, the appearance of any lithofacies $\mathrm{L_i}$ (W_k = $\mathrm{L_i}$) at any position k does not depend on underlying lithofacies located at positions k – 1, k – 2, *etc.* Therefore, the process which generated the sequence $\{W_k\}$ (*i.e.* sedimentation) does not “remember” its history. If the lithostratigraphic section is described with the matrix (2), such defined randomness is impossible: the process must ”remember” the previous layer in order to prevent the repetition of the same lithofacies in the following layer. Hence, such a case is called ”quasi-randomness”: if $W_k \neq W_{k\text{-}1}$, probabilities of the occurrence of all other lithofacies at position k are equal.

Discrete processes endowed with ”memory” extended back up to q steps are called the q-order Markov chains. If we assume that f_{ii} = 0, these are the so-called ”embedded chains” (Krumbein & Dacey, 1969). In practice (at least in sedimentological practice), the first-order chains are in use. Hence, the following considerations will be limited to such chains only.

RANDOMNESS TEST OF A LITHOSTRATIGRAPHIC SECTION

The idea of statistical verification of randomness or quasi-randomness hypotheses of a sequence $\{W_k\}$ *versus* alternative hypothesis that the studied sequence represents the first-order Markov chain is very simple. If the null-hypothesis is valid, the matrix of expected facies transitions $\boldsymbol{E}$ should be calculated and compared with the matrix of observed transitions $\boldsymbol{F}$. If the differences between both matrices are significant, the rejection of null-hypothesis is suggested. Hence, the crucial problem is the calculation of matrix $\boldsymbol{E}$ and the selection of proper test statistics.

The problems is well-known from the non-embedded chains (see *e.g.*, Anderson & Goodman, 1957). The elements of matrix $\boldsymbol{E}$ are given with the following formulae:

$$e_{ij} \quad \frac{n_i \; n_j}{n} \,, \qquad (3)$$

where

$$n_i \quad \underset{j}{\ } f_{ij} \;,\; n_j \quad \underset{i}{\ } f_{ij} \;,\; n \quad \underset{i,j}{\ } f_{ij}$$

and where the significance of difference between the matrices $\boldsymbol{F}$ and $\boldsymbol{E}$ can be verified with the statistics:

$$X^2 \quad \overset{m}{\underset{i,j}{\ }} \frac{(f_{ij} \quad e_{ij})^2}{e_{ij}} \,, \qquad (4)$$

which has the asymptotic distribution χ^2 with $(m – 1)^2$ degrees of freedom.

A more complicated problem arises when one considers successions described with the matrix (2), which have zero values along the main diagonal. In such a case, formula (3) is not applicable because it treats these elements equally with the others. Potter and Blakely (1968), who first paid attention to this problem and its significance in geological applications, simply proposed that zero elements of the matrix can be neglected. Consequently, the calculated value X^2 may represent the random variable of distribution χ^2 with $(m – 1)^2 – m$ degrees of freedom. Unfortunately, this concept is an oversimplification – after automatic removal of m elements from the matrix values e_{ij} will generally be underestimated; thus, we cannot expect that the statistics X^2 will meet the postulated distribution χ^2.

Read (1969) and Gingerich (1969) in their independently run studies attempted to overcome this weak point. According to these authors, elements of matrix $\boldsymbol{E}$ should be determined from the formula:

$$e_{ij} \quad \begin{matrix} \dfrac{n_i \; n_j}{n \quad n_i} \,, & i \quad j \\ 0 \qquad\quad , & i \quad j \end{matrix} \,, \qquad (5)$$

which guarantees the identical sums of rows in matrices $\boldsymbol{F}$ and $\boldsymbol{E}$ (although it does not guarantee the identity of columns sums). The calculated X^2 statistics should reveal the asymptotic distribution χ^2 with the degrees of freedom calculated as $(m – 1)^2 – m$ (after Read, 1969) or as $m(m – 2)$ (after Gingerich, 1969).

Unfortunately, also this concept is incorrect, as clearly demonstrated by Powers and Easterling (1982). Moreover, these authors defined the clue of the problem, *i.e.*, the lithostratigraphic sections containing the “forbidden” transitions of the same facies varieties are not independent by definition. In such cases the total “lack of memory” of depositional process does not occur. As a solution, Powers and Easterling (1982) applied the idea of quasi-independence of qualitative variables developed by Goodman (1968) and tested it with contigency tables. In such an attempt, the elements of matrix ***E*** are given by general formula:

$$e_{ij} \quad \begin{matrix} a_i b_j & , & i \quad j \\ 0 & , & i \quad j \end{matrix} , \qquad (6)$$

where values a_i and b_j are determined with relevant iterations.

Let m represent the number of rows/columns in the matrix of transitions frequency, n_{i+} be the sum of i row, n_{+j} be the sum of j column and ε be the required accuracy of iteration. Thus:

$$a_i^{(1)} \quad \frac{n_i}{(m \quad 1)} \quad , \quad i \quad 1,2,\ldots,m ;$$

$$b_j^{(1)} \quad \frac{n \quad j}{\underset{i \; j}{a_i^{(1)}}} \quad , \quad j \quad 1,2,\ldots,m ;$$

and

$$a_i^{(k)} \quad \frac{n_i}{\underset{j \; i}{b_j^{(k \quad 1)}}} \quad , \quad i \quad 1,2,\ldots,m ;$$

$$b_j^{(k)} \quad \frac{n \quad j}{\underset{i \; j}{a_i^{(k)}}} \quad , \quad j \quad 1,2,\ldots,m .$$

Iterations are run until the following condition is satisfied:

$$\left| a_i^{(k)} \quad a_i^{(k \quad 1)} \right| \quad , \quad i \quad 1,2,\ldots,m ,$$

$$\left| b_j^{(k)} \quad b_j^{(k \quad 1)} \right| \quad , \quad j \quad 1,2,\ldots m .$$

Another estimation method of expected numbers of transitions was presented by Davis (2002). In this procedure, the studied lithostratigraphic succession is treated as a ”censored” sample of a section, in which transitions between identical lithofacies are permitted and the matrix ***F*** for this sample differs from the observed matrix only with the elements located along the main diagonal. Under such assumption elements of matrix ***E*** can be estimated with iteration.

If m is the number of rows/columns in the matrix of transition frequency, f_{ij} is an element of this matrix, L is the preset initial value and l is the required accuracy of iteration, then:

$$e_{ij}^{(1)} \quad \begin{matrix} L & , & i \quad j \\ f_{ij} & , & i \quad j \end{matrix} ,$$

$$n_i^{(1)} \quad \overset{m}{\underset{j \; 1}{}} e_{ij}^{(1)} \quad , \quad n_{i \; 1}^{(1)} \quad \overset{m}{\underset{i \; 1}{}} n_i^{(1)} \quad ,$$

and

$$e_{ii}^{(k)} \quad \frac{n_i^{(k \quad 1)} n_i^{(k \quad 1)}}{n^{(k \quad 1)}} \quad , \quad i \quad 1,2,\ldots,m ;$$

$$n_i^{(k)} \quad \overset{m}{\underset{j \; 1}{}} e_{ij} \quad , \quad n^{(k)} \quad \overset{m}{\underset{i \; 1}{}} n_i^{(k)} .$$

Finally, if the following condition is met:

$$\left| e_{ii}^{(k)} \quad e_{ii}^{(k \quad 1)} \right| \quad , \quad i \quad 1,2,\ldots,m ,$$

we obtain:

$$e_{ij} \quad \begin{matrix} 0 & , & i \quad j \\ \dfrac{n_i^{(k)} n_i^{(k)}}{n^{(k)}} & , & i \quad j \end{matrix} .$$

Both the estimation methods of elements of matrix ***E*** are based upon iteration algorithms, the convergence of which has not been tested. Thus, the danger exists that in a particular case results of calculations will be unsatisfactory. The solution lies in the application of well-known Monte Carlo method, which enables the researcher to estimate directly the required values with the simulation of random process of sedimentation.

Let us assume that the studied lithostratigraphic section $\{W_k\}$ includes N layers, from which n_1 represents lithofacies L_1, n_2 represents lithofacies L_2, *etc.* If we randomly select layers in this section (having in mind that the succeeding layers cannot be lithologically identical) we will generate a new succession (”section”), which will meet the requirement of quasi-independence. If we generate a large number (*e.g.*, 1,000) of such sections we can calculate for each the matrix of the numbers of facies transitions. After averaging the matrices, we will obtain the estimation of the expected numbers of facies transitions e_{ij}.

The advantages of the proposed method are: the absence of any initial assumptions (except for repeatability of lithofacial varieties) and the independence of the results on iteration procedures. Calculations can be run with any available generator of pseudo-random numbers, e.g., that operating in commonly used *Excel* software.

It is obvious that, despite selected estimation method of the elements of matrix ***E***, the final step of randomness test of deposition process recordered in a lithostratigraphic section is the calculation of X^2 statistics (see formula 6).

COMPARISON OF ESTIMATION METHODS OF EXPECTED NUMBERS OF FACIES TRANSITIONS

Below, an application is presented of estimation methods of the expected numbers of facies transitions. True data originate from the HEBCH-1 succession, which represents the paralic series of the Upper Silesian Coal Basin. The succession includes 126 layers of eight lithological varieties: claystones (SH – 47 layers), coaly claystones and coaly shales (CS – 3 layers), coal (COAL – 33 layers), fine-grained sandstones (SF – 20 layers), mudstones (MU – 17 layers), medium-grained sandstones (SM – 2 layers), coarse-grained sandstones (SC – 1 layer) and heteroliths, *i.e.* rocks composed of the sets of clay-sandstone-mudstone laminae (HE – 3 layers). The observed matrix ***F*** of transitions between these facies varieties is shown below.

	SH	CS	COAL	SF	MU	SM	SC	HE	Σ
SH	0	0	23	11	10	1	1	0	46
CS	2	0	1	0	0	0	0	0	3
COAL	26	3	0	4	0	0	0	0	33
SF	9	0	4	0	6	0	0	1	20
MU	9	0	3	3	0	0	0	2	17
SM	1	0	0	1	0	0	0	0	2
SC	0	0	0	0	0	1	0	0	1
HE	0	0	2	0	1	0	0	0	3
Σ	47	3	33	19	17	2	1	3	125

Below, matrices are shown of the expected numbers of transitions in a quasi-random column calculated with the three methods:

Powers-Easterling method

	SH	CS	COAL	SF	MU	SM	SC	HE	Σ
SH	0.00	1.52	21.31	10.75	9.40	1.00	0.50	1.52	46.00
CS	1.53	0.00	0.70	0.35	0.31	0.03	0.02	0.05	2.99
COAL	21.54	0.70	0.00	4.99	4.36	0.47	0.23	0.70	32.99
SF	11.37	0.37	5.22	0.00	2.30	0.25	0.12	0.37	20.00
MU	9.51	0.31	4.36	2.20	0.00	0.21	0.10	0.31	17.00
SM	1.02	0.03	0.47	0.24	0.21	0.00	0.01	0.03	2.01
SC	0.50	0.02	0.23	0.12	0.10	0.01	0.00	0.02	1.00
HE	1.53	0.05	0.70	0.35	0.31	0.03	0.02	0.00	2.99
Σ	47.00	3.00	32.99	19.00	16.99	2.00	1.00	3.00	124.98

Davis method

	SH	CS	COAL	SF	MU	SM	SC	HE	Σ
SH	0.00	1.44	21.16	11.06	9.14	0.96	0.48	1.44	45.68
CS	1.44	0.00	0.72	0.38	0.31	0.03	0.02	0.05	2.95
COAL	21.16	0.72	0.00	5.53	4.57	0.48	0.24	0.72	33.42
SF	11.06	0.38	5.53	0.00	2.39	0.25	0.13	0.38	20.12
MU	9.14	0.31	4.57	2.39	0.00	0.21	0.10	0.31	17.03
SM	0.96	0.03	0.48	0.25	0.21	0.00	0.01	0.03	1.97
SC	0.48	0.02	0.24	0.13	0.10	0.01	0.00	0.02	1.00
HE	1.44	0.05	0.72	0.38	0.31	0.03	0.02	0.00	2.95
Σ	45.68	2.95	33.42	20.12	17.03	1.97	1.00	2.95	125.12

Monte Carlo method (for 1000 repetitions)

	SH	CS	COAL	SF	MU	SM	SC	HE	Σ
SH	0.00	1.56	21.68	10.49	9.17	1.01	0.63	1.46	46.00
CS	1.45	0.00	0.68	0.41	0.36	0.03	0.02	0.05	3.00
COAL	21.45	0.67	0.00	5.11	4.41	0.48	0.17	0.72	33.01
SF	11.51	0.38	5.06	0.00	2.35	0.24	0.09	0.38	20.01
MU	9.62	0.30	4.24	2.22	0.00	0.20	0.08	0.33	16.99
SM	0.95	0.03	0.47	0.26	0.24	0.00	0.01	0.04	2.00
SC	0.54	0.01	0.17	0.13	0.13	0.01	0.00	0.01	1.00
HE	1.48	0.05	0.70	0.38	0.34	0.03	0.01	0.00	2.99
Σ	47.00	3.00	33.00	19.00	17.00	2.00	1.01	2.99	125.00

SUMMARY

The examples presented above demonstrate that all three applied methods produce almost identical results; hence, all provide the same conclusions after the randomness test was calculated. It must be emphasized, however, that the Davis method produced perfectly symmetric matrix of transition numbers, which means that this method neglects the difference in sums of rows and columns for the first and last element. This difference can be important for short lithostratigraphic sections, which, unfortunately, are quite common in geological practice. For such sections, a much better solution is the application of the Powers-Easterling or Monte Carlo methods. However, for long wells and when the Monte Carlo method is used, the length of the studied sequence is unimportant.

The problem of estimation of expected values is so critical because sedimentological conclusions are based upon differences between expected and observed numbers of facies transitions in the studied lithostratigraphic sections. Hence, incorrect estimations of expected numbers of transitions will inevitably lead to erroneous differences and, consequently, will generate wrong genetic conclusions.

REFERENCES

Anderson, T.W. & Goodman, L.A., 1957. Statistical inference about Markov chains. *Annals of Mathematical Statistics*, 28(1): 89-110.

Carr, T. R., Horowitz, A., Hrabar, S. V., Ridge, K. F., Rooney, R., Straw, W. T., Webb, W. & Potter, P.E., 1966. Stratigraphic sections, bedding sequences and random processes. *Science*, 154: 1162-1164.

Davis, J.C., 2002. *Statistics and Data Analysis in Geology.* Wiley & Sons, New York, 678pp.

Gingerich, P. D., 1969. Markov analysis of cyclic alluvial sediments. *Journal of Sedimentary Petrology*, 39(1): 330-332.

Goodman, L. A., 1968. The analysis of cross-classified data. *Journal of the American Statistical Association*, 63(324): 1091-1131.

Krumbein, W.C., 1967. *FORTRAN IV computer programs for Markov chain experiments in geology.* Kansas Geol. Survey Comp. Contr., 13, 38pp.

Krumbein, W. C. & Dacey, M.F., 1969. Markov chains and embedded Markov chains in geology. *Journal of the International Association for Mathematical Geology*, 1(1): 79-96.

Potter, P.E. & Blakely, R.G., 1967. Generation of a synthetic vertical profile of a fluvial sandstone body. *Society of Petroleum Engineers Journal*, 7: 243-251.

Potter, P.E. & Blakely, R.G., 1968. Random processes and lithologic transitions. *Journal of Geology*, 76(2): 154-170.

Powers, D. W. & Easterling, R.G., 1982. Improved methodology for using embedded Markov chains to describe cyclical sediments. *Journal of Sedimentary Petrology*, 52(3): 913-923.

Read, W. A., 1969. Analysis and simulation of Namurian sediments in central Scotland using a Markov-process model. *Journal of the International Association for Mathematical Geology*, 1(2): 199-219.

Vistelius, A.B. & Faas, A.V., 1965. The mode of alteration of strata in certain sedimentary rock sections. (In Russian, English summary). *Doklady Akademii Nauk SSSR*, 164: 40-42.

Vistelius, A.B. & Feygelson, T.S., 1965. Stratification theory. (In Russian, English summary). *Doklady Akademii Nauk SSSR*, 164: 20-22.