**Conceptual Framework of Finite Markov Chain Stochastic Processes and Their Application to Representative Geological Case Studies.**

**ABSTRACT**

Stratigraphy is a key concept in understanding the depositional domain of a sedimentary unit and its prospects for the exploration of coal and hydrocarbon. Generally, stratigraphy analysis uses a convenient qualitative approach to interpret sedimentary environments and models are reconstructed based on visual description is well received among sedimentological researchers to conduct depositional and lithofacies analysis. The interpretations and environmental models derived from such generally subjective data are therefore wanting. New statistical tools in stratigraphy including semi-quantitative and quantitative models for categorical data analysis were introduced to compliment subjective stratigraphic analysis. These quantitative methods based on a simple probability theory are used to predict the successive units and corresponding process in a sedimentary sequence. The Finite Markovian statistics of vertical and lateral variability were made possible by integrating Walther’s Law, which states that the vertical succession of sedimentary facies in a geological column reflects the lateral changes and succession of depositional environments, thus simulating both lateral and vertical facies variability. Minor erosional breaks are regarded as probabilistic within the model. Based on stochastic methods it is observed that the early Permian coal bearing cycles are auto-cyclic in nature. Coal measure cyclothems or fluvial fining-upward cycles around the world are good examples of sedimentary succession laid down under the control of Markovian process.

The present contribution aims to highlight the scope of stochastic processes-- including Finite Markov Chain, Quasi-independence, Markov Reversibility and Marginal homogeneity. These quantitative methods enables more predictability in analysis of stratigraphic patterns and quantify sedimentary cycles for proper correlation within and between oil bearing and coal bearing successions and it’s bearing on exploration of coal and hydrocarbon and their development. The data used in this work comes from the vertical stratigraphic outcrop sections/subsurface borehole logs from the early Permian Gondwana stratigraphy of peninsular India.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Key Words:** StochasticFinite Markov Chain, Quasi-independence, Reversible Markov Chain, Coal Bearing Sequences, Lithofacies Analysis.

**1. INTRODUCTION**

Sedimentologists in their field work usually provide records of sedimentary strata as a vertical section of successive lithofacies. Much of the value of stratigraphy focuses on its record of change. Processes such as sea level cycles, uplift, and climate change are sought from stratigraphy to write the story of the past or, inversely, to explain shifts in the stratigraphic patterns observed [1]. The key to the interpretation of the sedimentary record is to study facies association, in particular their vertical and lateral position [2]. Vertical variations of lithofacies within a given sequence play an important role in the recognition of depositional environment and their lateral dispersal. In terms of basic theory, notable developments have taken place that encompass refinements in geological and geophysical concepts including several enhancements to sequence-stratigraphic and facies distribution models, better knowledge of seismic response3 mechanism to small-scale stratigraphic structures and different rock types, as well as the emergence of new technologies, such as the one allows the restoration of original depositional architecture and spectral decomposition [3, 4]. It is of particular importance in the light of the widely accepted law of the correlation of facies as proposed by Walther and elaborated by [5] and more recently by [6] According to this law, only those facies can be superimposed which can be observed beside each other at a given time. Importantly, only a gradual transition from one facies to another implies that the two facies represent environments that adjacent laterally. Comparison of lithofacies from the study area with standard facies models for different depositional systems is a useful practice and adopted by several researchers. However, reliance on model-driven interpretations often leads to the imposition of perceived order, even when natural stratigraphic patterns may lack systematic arrangement [7]. Traditionally, reconstruction of sedimentary environments relies on visual descriptions of facies and lithologic variables, which are inherently subjective. Consequently interpretations and environmental models derived from subjective data are often limited in reliability. New tools in stratigraphy including semi-quantitative and quantitative techniques for data analysis were introduced to complement descriptive stratigraphic analysis. Techniques such as the finite Markov chain analysis, the time series analysis, power spectral density analysis and Kolmogorov functions introduce and precision in analyzing and interpretation of stratigraphic models and depositional environments and reduce subjective impact on interpretation of the sedimentological record. They also provide multiple working hypotheses to detect and define stratigraphic trends sequences. Almost all use the probability matrices and employ the principles of Finite Markov chains statistics. This method was developed because of the urgent need to a better understanding how the natural process interacts and produces the end product we record on the litho-stratigraphic sections. The Finite markov chains are increasingly considered standard tools in sedimentological analysis, but still many publications whose authors do not fully understand the applied stochastic statistical methods and, thus their conclusions are not fully correct. Therefore, it seems reasonable to present once again the principle of Finite Markov model applications to the analysis of litho-stratigraphic successions.

The present research has two-- fold objectives; 1) the main purpose is to provide an explicit treatment of the mathematical foundations on which application of Markov stochastic process in geology based, but not readily available in publications normally used by geologists and 2) the effectiveness of the Markov Chain to predict and quantifies sedimentary cycles while dealing with the geological interpretation of juxtaposition tendencies of facies in vertical stratigraphic successions is evaluated by the interval on borehole data. The results show that the proposed stochastic methods have good performance and robustness.

**2. MATERIALS AND QUANTITATIVE METHOD**

The basic requirement for such study is the vertical lithological transition data. The Early Permian Gondwana stratigraphy of Peninsular India is selected for data collection because of the availability of borehole logs drilled extensively by the Government and semi-government agencies for coal exploration. The length of borehole logs varies between 100-210m. This apart, the small (5-10m) measured outcrop sections are also used to supplement the borehole transitions. The transition data used in this study from (i) Karharbari Formation, Giridih sub-basin (Koel-Damodar valley), (ii) Karharbari Formation, Talchir sub-basin (Son-Mahanadi valley), (iii) Barakar Formation (Pranhita-Godavari valley), and (iv) Barakar Formation, Saharjuri sub basin (Koel-Damodar valley) of Gondwana Master Basin of Peninsular India.

The Early Permian Gondwana stratigraphy (Karharbari and Barakar formations) of Peninsular India consists of an interbedded sequence of coarse to medium grained sandstone, arenaceous and argillaceous shale, carbonaceous shale and coal with variable proportions in space and time [8]. Compositional studies based on large number of borehole logs, quarry sections and accessible outcrops have shown that the Karharbari Formation abounds in sandstone (≈ 70-80%), with associated shale (≈8-12%) and coal (5-16%). The overlying Barakar Formation shows a reduction in bulk sandstone (≈32 - 62%), and increase in shale (≈20-62%) and coal and shaly coal (≈6- 42%). Eight lithofacies are identified in the given succession from the borehole log and outcrop/quarry sections as:(i) very coarse to sandstone, occasionally pebbly (ii) coarse to very coarse sandstone, (iii) medium to coarse sandstone, (iv) fine to medium sandstone, (v) ripple drift laminated fine sandstone, (vi) interbedded fine grained sandstone and argillaceous shale, (vii) arenaceous and argillaceous shale and (viii) interbedded carbonaceous shale, shaly coal, and coal. Recognition of changes between lithologies will not be difficult as it is necessary to simplify stratigraphic sections down to an easily manageable number of gross litho-types (lithological state) in order to prevent transition tendencies from being too diffused throughout the count matrix. The above lithofacies are condensed to only five lithofacies which are distinctly marked in borehole as well as in outcrop sections. It would be possible to divide lithologies into a larger number of states by subdividing litho- types on the basis of color, presence or absence of cross bedding and other such criteria. This would largely set aside the problem of multistory lithologies, but require much greater quantity of observational data for statistically meaningful relationships to appear. The condensation of lithologies does not affect the overall interpretation of facies model inasmuch as help to obtain sizeable number of data of lithologic transitions. These five facies are:

* Facies (COSD): coarse to very coarse grained sandstone,
* Facies (MFSD): medium to coarse grained sandstone,
* Facies (FSD): fine grained to ripple laminated fine sandstone
* Facies (SH): argillaceous shale and interbedded fine sandstone and shale,
* Facies (C): coal and shaly coal.

All these lithofacies have been very well interpreted for depositional processes, wherein coarse grade facies (COSD and MFSD) correspond to in-channel/point bars, fine grade (FSD and SH) to overbank and coal (C) to back swamps of fluvial system [8, 9,and 10]. These lithofacies are used to structure separately (1) continuous time and (2) discrete Markov models.

**3. FINITE MARKOV CHAIN ANALYSIS**

Finite Markov chain is a class of probabilistic model created by Andrey Markov that outlines the probability associated with a sequence of events occurring based on the state in the previous event, which means that the event will not be affected by how current situation comes. In simple words the probability that *(n+1)* th events will be depends only on the *(n)* th events not the complete sequence of events that came before *n*. This property is known as Markov Property. It’s very common and easy to understand due to their versatility have achieved widespread application in the quantitative sciences for modeling stochastic processes. The main purpose of the Markov Stochastic process is to specify the probability of transitioning from one to another succeeding event. In sedimentary framework, Finite Markov process corresponds to successive pulses of deposition which are random and dependent in their occurrence, so that past deposition will influence present and future deposition. The “memory” of the depositional process may be short or long, and when it extends backward only one step, it is described as first-order Finite Markov process. Recent studies show that Finite Markov process with a one step memory make better models for the analysis of sedimentary successions than models without a memory [11] . The Markov model therefore, assumes that a lithofacies state is influenced by that of the underlying lithofacies. The literature on first- order Finite Markov chain is now rapidly growing and any test cases have been published in the last decade in facies analysis of various types of deposits by [12, 10, and 13] and many more.

The starting point in Finite Markov chain analysis is the transition count matrix. This is a two-dimensional array which tabulates the number of times that all possible vertical lithologic transitions occur in a given stratigraphic sequence. The lower bed (underlying) of each transition couplet is given by the row number of the matrix, and the upper bed (overlying) by the column number, each lithofacies being assigned a code number for the purpose of the analysis. An example is given in Table 1, in which the lithofacies codes are as follows: A. Sandstone: B. Shale: C. Coal.

Elements in the transition count matrix are hereafter referred by the symbol *fij*, where i= row number and j = column number. It will be noted that where i=j zeros are present in the matrix, i.e. transitions have only been recorded where the lithofacies shows an abrupt change in character, regardless of the thickness of the individual bed.

[11] Described three models of structuring lithofacies states observed in field as well as in borehole logs for finite Markov chain analysis for simulating sedimentary cycles. These are;

(i). The regular Markov where stratigraphic intervals are sampled at fixed or regular vertical intervals of equal thickness to generate the data for simulating cycles designated as continuous time Markov chain model (CTMC) and does not emphasize changes in the depositional environment. It correspond to [12] fixed interval Markov chain;

(ii) The embedded or discrete time Markov chain (DTMC) matrix which emphasize every lithological change and a good method for understanding the evolution of depositional environments and processes. A lithologic state can pass upwards into similar lithology resulting non-zero elements in diagonal of tally count matrix (i.e. Nij ≠0). It correspond to the “embedded Markov chain “of [13] and;

(iii) The multistory Markov matrix which is a variant of the embedded or discrete Markov matrix in which similar lithofacies but with different texture or sedimentary structures are counted for structuring of a tally count matrix for lithological transitions. Since a transition is supposed to occur only when it results in a different lithology, the diagonal elements are all zeros in the resulting tally count matrix (i.e. *nij* =0). The discrete as well as continuous multistory time Markov model is applied in this study.

**3.1 Continuous Time Markov Chain (CTMC)**

In the past, various researchers [7, 17, and18], have used the theory of discrete time Markov chain (DTMC) to describe the structure of a sequence of lithologies, whereas very few workers [19, 20, and 21] have been carried out for Markov process for stratigraphic simulation using continuous time Markov chain parameters. Markov process with a continuous time (CTMC) parameters are more satisfactory for describing sedimentation than discrete time Markov chain (DTMC) because they treat sedimentation as a natural process that happens continuously (i.e., which is unbroken in time). No doubt they yield not only the same information as a discrete time analysis, but also give information about the distribution of the thicknesses of the lithologies. However, the simulated stratigraphy using continuous time Markov model helps in predicting subsurface deposits specially coal and hydrocarbons [22]. The sole object of such model is to simulate stratigraphy for prediction of associated deposits. Three distinct lithofacies states of sandstone (A), shale (B) and coal (C) from Karharbari Formation of Giridih sub basin are used in this study. The first step in continuous time model is to compute *Transition Count Matrix (f ij),* where *i* and *j* corresponds to row and column numbers. Trial and error suggested 2.4 m thickness interval, and transitions are counted continuously from lithologic succession. Such matrices provide idea of relative thicknesses of facies, and diagonal of the matrix contains non-zero entries. The other matrix of CTMC is named as *Transition Probability Matrix (p ij*). It records the probability of actual vertical transitions, and is computed by dividing each cell value (*f ij)* by corresponding row total *(n+).* These probability values are then converted to *Fixed Probability Vector* by 19 successive iterations where transition probability in each attains similar non-zero positive values sum to unity. It represents the **‘***Equilibrium Stage’*of Markov chain where the cell values corresponds relative percent thickness of each facies [15]. The transition probability matrix (*p ij)* is used to simulate Karharbari stratigraphy. To begin with, the original probability values are converted to cumulative probability by successive addition of first and second *p ij* value, and then first, second and third *pij* elements to obtain second and third *pij* values of the same row of *Cumulative**Probability Matrix*(Table 1). The cumulative values are used as input to select appropriate transitions by way of random numbers for simulation experiments [19].

**Table 1.** **Continuous Time Markov Chain Matrices , Karharbari Formation, Giridih sub-basin.**

(A).Transition Count Matrix *(fij)*

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  lithofacies |  Sandstone (A) |  Shale (B) |  Coal (C) | row total *(n+i)* |
|  Sandstone (A) |  256 |  35 |  07 | 298 |
|  Shale (B) |  30 |  25 |  09 | 64 |
|  Coal (C) |  12 |  04 |  21 | 37 |
|  (n+j) |  298 |  64 |  37 | *n++* = 399 |

(B).Transition Probability Matrix *(pij)*

|  |  |  |  |
| --- | --- | --- | --- |
|  lithofacies |  Sandstone (A) |  Shale (B) |  Coal (C) |
|  Sandstone (A) |  0.859 |  0.118 |  0.022 |
|  Shale (B) |  0.462 |  0.400 |  0.138 |
|  Coal (C ) |  0.225 |  0.085 |  0.660 |

(C).Fixed Probability Matrix after 19 successive steps

|  |  |  |  |
| --- | --- | --- | --- |
|  lithofacies |  Sandstone (A) |  Shale (B) |  Coal (C) |
|  Sandstone (A) |  0.727 |  0.159 |  0.113 |
|  Shale (B) |  0.727 |  0.159 |  0.113 |
|  Coal (C)  |  0.727 |  0.159 |  0.113 |

(D).Cumulative Probability Matrix

|  |  |  |  |
| --- | --- | --- | --- |
|  lithofacies |  Sandstone (A) |  Shale (B) |  Coal (C) |
|  Sandstone (A) |  0.859 |  0.977 |  1.000 |
|  Shale (B) |  0.859 |  0.977 |  1.000 |
|  Coal (C ) |  0.859 |  0.977 |  1.000 |

Higher transition probability from sandstone to sandstone implies thick multistory sandstone bodies as observed in field. The constant values of sandstone, shale and coal in fixed probability vector representing their relative percent (72.7%, 15.9%, and 11.3%) compare very well with the field observations. Figure 1 illustrates the actual and simulated Karharbari Formation along with the transition count and transition probability matrices, which are fairly comparable. The simulated succession may, therefore, be used to predict subsurface stratigraphy and associated coal deposits.



**Figure 1.** **Comparative actual and simulated stratigraphic sections of the Karharbari Formation of Giridih sub-basin. Note the similarity in Transition count and Transition probability matrices of the two sections.**

**3.2. Discrete time Markov Chain (DTMC)**

**3.2.1. The Transition Count Matrix (TCM)**. The starting point in this method was to structure the Transitions Count Matrix (TCM). This is a two-dimensional array which tabulates the number of times all possible vertical lithofacies transitions occur in any given stratigraphic succession that has been subdivided into its component facies (Table 2).In geological terms, the table indicates the number of times each facies is overlain by another. The lower bed of each transition couplet is given by the row number of the matrix and denotes a definite lithology, and columns give the composition of stratum resting immediately above a chosen lithology. Hence, if there were *n* lithofacies then the total of both the row and column is *n*-1.Elements in the transition Count Matrix (TCM) referred by the symbol *nij* where *i* = row number and *j* = column number. It will be noted that where *i =j* the leading diagonal of the transition matrix has zeros elements in the matrix, i.e. transitions have only been recorded where the lithofacies shows an abrupt change in character, regardless of the thickness of the individual bed. This means that once a lithofacies state has been entered, the next state must be different; no return to the same lithofacies state is possible. The transition count matrix (TCM) is also called tally matrix [15] and summing the individual tallies *nij*either rows or columns must give the total number of transitions.

**3.2.2. The Probability Matrices.** From the transition count matrix two probability matrices have been derived. The first in an independent trials probability matrix (ITPM) composed on *r ij* which represents the probability of the given transition occurring randomly. Given any lithofacies state *i* the probability of this

state being succeeded by any other lithofacies state *j* is dependent on the relative proportions of the various states present. The value in each cell of this matrix was calculated as:

 *rij = f+j / (n++ ─ f+i)*

Where *r ij =* random probability of transition from state *i to j*, n++ = total number of transitions for all lithofacies state of rows and columns used, *f+j* = random number of occurrences of facies j (i.e. column total for facies *j*) and *n+i =* number of occurrences of lithofacies state *i.*

In the random probability matrix, all the diagonal cells are ‘structurally empty’, because of the impossibility of recognizing the transition of a given lithofacies upward into another bed of the same lithofacies.

A second matrix, containing elements *pij* was constructed by converting the observed number of transitions to actual transition probabilities for vertical sampling interval.This was done by dividing the number of transitions in each cell by the row total as follows:

 *p i j= f i j / n i+*

The values of *pij* matrix sum to unity along each row and they will necessarily reflect the presence of any Markovian dependency relationship. [23] Suggested the use of a difference matrix which is produced when the matrix showing expected transition probability (*r ij*) is subtracted from the transition probability matrix (*p ij )* of observed transitions.

 *d ij = p ij – r ij*

Positive (+) entries of the difference matrix serve to emphasize the preferred cyclic sequence or Markov property in the given succession by indicating which transitions have occurred with greater than random frequency. It is, nevertheless, desirable to test statistically whether the difference between *p ij* matrix and *r ij* matrix are due to random chance. A chi-square test is suitable for this purpose as suggested by [15, 24] have been applied by many researchers to test for the Markovian property in different sedimentary successions [12,10] To visualize the transition patterns, a Facies Relationship Diagram (FRD) is constructed to represent lithofacies transition path using the positive entries in the difference matrix.

**4. Test of Significance (Chi-square test)**

 Differences observed between the Transition Probability Matrix (TPM) and the independent Trails Probability Matrix (ITPM) (Table 2) may appear considerable. However, such variations might also result from random chance. Therefore, a statistical test is necessary to determine whether the observed differences reflect true dependence among facies or are due to stochastic variation. A chi-square test (ϰ2) is suitable for this purpose to test the dependency between any two facies states, during constructing the facies transitions in sequence. A chi-square statistics is given by [23, 24]

 *ϰ2 = ∑I j (fij – fi rrj )2 / fi rij*

has asymptotically the chi-square distribution with *(n-1)2-n* degree of freedom for *n x n* facies transition matrix. The null hypothesis is that the vertical succession of strata was derived by random variation in the depositional mechanism.

**5. Entropy of Discrete Time Markov Chain**

The starting point in an entropy analysis of discrete Markov chain is a tabulation of transition matrix between vertical lithologic states observed in each given stratigraphic succession (Table 3). Let *f i j* denotes the number of upward transitions from lithologic state *i* to *j*, row totals *n i+*, column totals *n +j* and then grand total *n++* is defined as

 *n++* = *∑ i ∑ j f ij*

The upward (forward) transition probability of *i* ***→*** *j* associated with element *f ij* is *p ij* = *f ij*/ *n i+*, which is equated with the ‘entropy after deposition [E (post)]’ (i.e. across the row) has been calculated with respect to *i* using relationship E (post) = *∑i* *p ij log2 p ij,* where *n* is relative frequency that lithofacies state *j* follows *i*  and *n* is number of lithologic state.

The second matrix, containing *q ji* which represent the probability of the giving transition being preceded by and other transition i.e. downward (backward) transition probability of *j → i* associated with element *f ij* is *q j i* = *f i j* / *n +j* ; which is equated with the ‘entropy before deposition [E (pre)]’ can also be calculated in a resemble manner E (pre) = *∑i* *q ji log2 q ji* where E (pre) is entropy before deposition (i.e. along the column) with respect to lithologic state *i*; *q*ji isrelative frequencythat state *j* precedes state *i*.

[E (pre)] and [E (post)] indicate the variety of lithological transitions which immediately lead into, and follow from, that lithologic state and they were as a useful supplement to the transition probability information. By plotting [E (pre)] against [E (post)] for each lithological state one can make some interpretation of the style of cyclicity and the way in which cycles are deposited. [25] shows a number of diagrams of the distribution of [E (pre)] versus [E (post)] for idealized, truncated symmetrical (ABCDCBA) and asymmetrical (ABCDABCD) lithological succession.

Apart from entropies with respect to individual sets, maximum entropy possible in a system where *n* lithological states operate has been calculated as *E (max) = -log2 1/ (n-1)*. Similarly the entropy of the whole sedimentation unit can be calculated as

E (system) = *∑ i ∑ j* *r i j log2 r i j* , where *r i j*  = *f i j* / *n ++* : where *f i j* are entries in the tally count matrix and *n++* denoted total number of lithological states, which can be used for deciphering the overall depositional environment of cyclical sequences and takes a value between - log2 1/*n* and – log2 1/*n(n-1)* where *n* is the number of lithological states. A plot of this statistics against the number of lithological states in the system can prove to be useful indicator of the overall environment of deposition of cyclical successions. The calculated entropy values for the whole sedimentation vary within 4.322-2.321 in this study, indicating possibly fluctuating/ oscillatory nature of the depositional system and dominance of non-random events in the Karharbari succession.

The various probability matrices for discrete time Markov chain (DTMC) that can be calculated for the example given below is given in Table 2 and 3 and those for entropies matrices is given in Table 4 will be discussed in the section on result.

**6. Binomial Probability**

A positive (+) value in the difference matrix (i.e., pij - *rij* > 0 ) as discussed elsewhere doesn’t prove that the difference is statistically significant. One limitation of the previous methods is the inability to determine which transitions are truly significant as outlined by [26].

To address this [27] proposed the use of binomial probability of at *N obs* successes in *N* trails, this corresponds to the binomial probability (BP), and is given by

 ∑ n=N = C (N, n) p n q obs

 n=nobs

where *C (N, n)* = the number of possible combinations of *N* states taken *n* at a time, and is given by *:* C(N, n) = N! / (N, n)! n!, *N =* total number of upward transitions of any lithofacies into all other lithofacies i.e. the row totals in transition count matrix , n = number of upward transitions of any lithofacies into any other lithofacies, *q = 1-p, p* = probability of success on a single trail and *q* = probability of failure on a single trails.

A null hypothesis (H0) was set up; all transitions of any facies into another facies occur randomly succession. The independent N trails are accepted at significance level of 90%. A criterion is set up to accept or reject the null hypothesis (H0) if the value of computed binomial probability (BP) is greater than or equal to the level of significance chosen- it will be rejected otherwise the null hypothesis would be accepted. Table 2E shows the summary of binomial probability for all positive difference with a level of significance of 0.10 and 0.05%.

**6.1 Application and Geological interpretation**

Tally count data for the 21 borehole logs were added and one-step embedded bulk tally count matrix (*f ij*) was compiled. Table 1 records the data for the transition count matrix and separate independent trails (r*ij)*, transition probability (*p ij*) and difference (*d ij*) matrices. To determine if transition pattern is a product of chance events (i.e. random) or the sedimentation process has a ‘memory function’ (i.e. Markovian), a test of significance was applied to the result. The chi-square statistics after [24] have been applied for this purpose as have been used by several workers [28, 29 and 30]. In this test, a null hypothesis, Ho is proposed which assumes that the transition pattern generalized by the sedimentary process was random event i.e. non Markovian. The test ϰ2 is applied to reject or uphold this hypothesis.

**Table2**. **Markov matrices, Binomial probability (BP) and Chi-square statistics (ϰ2) of the Karharbari Formation, Talchir Coalfield.**

(A).Transition Count Matrix (*f ij*)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  COSD |  MFSD |  FSD |  SH |  C |  n+i |
|  COSD |  0 |  41 |  47 |  13 |  02 |  103 |
|  MFSD |  32 |  0 |  38 |  14 |  05 |  89 |
|  FSD |  41 |  21 |  0 |  48 |  11 |  121 |
|  SH |  15 |  16 |  26 |  0 |  43 |  100 |
|  C |  17 |  09 |  09 |  26 |  0 |  61 |
|  n+j |  105 |  85 |  120 |  101 |  61 | *n*++= 474 |

(B).Independent Trial Matrix (*r ij*)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  COSD |  MFSD |  FSD |  SH |  C |
|  COSD |  0 |  0.240 |  0.326 |  0.269 |  0.164 |
|  MFSD |  0.267 |  0 |  0.314 |  0.259 |  0.158 |
|  FSD |  0.291 |  0.252 |  0 |  0.283 |  0.172 |
|  SH |  0.275 |  0.237 |  0.323 |  0 |  0.163 |
|  C |  0.249 |  0.215 |  0.293 |  0.242 |  0 |

C).Transition Probability Matrix (pij)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  COSD |  MFSD |  FSD |  SH |  C |
|  COSD |  0 |  0.398 |  0.456 |  0.126 |  0.019 |
|  MFSD |  0.359 |  0 |  0.427 |  0.157 |  0.056 |
|  FSD |  0.338 |  0.173 |  0 |  0.396 |  0.091 |
|  SH |  0.150 |  0.160 |  0.260 |  0 |  0.430 |
|  C |  0.278 |  0.147 |  0.147 |  0.426 |  0 |

(D). Difference Matrix (*dij*)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  COSD |  MFSD |  FSD |  SH |  C |
|  COSD |  0 |  **+0.138** |  **+0.130** |  -0.143 |  -0.145 |
|  MFSD |  **+0.092** |  0 |  **+0.196** |  -0.102 |  - 0.001 |
|  FSD |  **+0.047** |  -0.079 |  0 |  **+0.113** |  -0.081 |
|  SH |  - 0.125 |  -0.077 |  -0.063 |  0 |  **+0.267** |
|  C |  **+0.028** |  -0.068 |  -0.146 |  **+0.184** |  0 |

Chi-square Test: ϰ2 (cal) = 61.34 for 15 degrees of freedom; limiting value 32.80 at 99.5% confidence level.

**(E).Binomial probability (BP)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Transition |   *p* |  *q* |  N |  n | Probability |
| COSD→MFSD |  0.398 |  0.602 |  103 |  41 |  0.080 |
| MFSD→ FSD |  0.456 |  0.544 |  103 |  47  |  0.078 |
| MFSD→COSD |  0.359 |  0.641 |  89 |  32 |  0.089 |
| MFSD→FSD |  0.427 |  0.573 |  89 |  38 |  0.085 |
| FSD→COSD |  0.338 |  0.662 |  121 |  41 |  0.076 |
| FSDSH |  0.396 |  0.604 |  121 |  48 |  0.073 |
| SH→C |  0.430 |  0.570 |  100 |  43 |  0.080 |
| C→COSD |  0.278 |  0.722 |  61 |  17 |  1.746x10-15 |
| C→SH |  0.426 |  0.574 |  61 |  26 |  1.467x10-15 |

**Table3. Transition Count (*f ij*), Upward transition (*pij*), Independent trail (*rij*) matrices and Entropy values for the Karharbari Formation, Talchir Coalfield.**

(A).Transition Count Matrix (*f ij*)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  COSD |  MFSD |  FSD |  SH |  C |  n+i |
|  COSD |  0 |  41 |  47 |  13 |  02 |  103 |
|  MFSD |  32 |  0 |  38 |  14 |  05 |  89 |
|  FSD |  41 |  21 |  0 |  48 |  11 |  121 |
|  SH |  15 |  16 |  26 |  0 |  43 |  100 |
|  C |  17 |  09 |  09 |  26 |  0 |  61 |
|  n+j |  105 |  85 |  120 |  101 |  61 | *n*++= 474 |

(B).Upward Transition Probability Matrix (*pij*)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  COSD |  MFSD |  FSD |  SH |  C |
|  COSD |  0 |  0.398 |  0.456 |  0.126 |  0.019 |
|  MFSD |  0.359 |  0 |  0.427 |  0.157 |  0.056 |
|  FSD |  0.338 |  0.173 |  0 |  0.396 |  0.091 |
|  SH |  0.150 |  0.160 |  0.260 |  0 |  0.430 |
|  C |  0.278 |  0.147 |  0.147 |  0.426 |  0 |

(C).Downward Transition Probability Matrix (*qji*)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  COSD |  MFSD |  FSD |  SH |  C |
|  COSD |  0 |  0.471 |  0.391 |  0.126 |  0.032 |
|  MFSD |  0.305 |  0 |  0.316 |  0.136 |  0.082 |
|  FSD |  0.390 |  0.241 |  0 |  0.466 |  0.180 |
|  SH |  0.143 |  0.183 |  0.216 |  0 |  0.705 |
|  C |  0.162 |  0.104 |  0.075 |  0.252 |  0 |

(D). Independent trails Matrix (rij)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  |  COSD |  MFSD |  FSD |  SH |  C |
|  COSD |  0 |  0.086 |  0.099 |  0.027 |  0.004 |
|  MFSD |  0.067 |  0 |  0.080 |  0.029 |  0.010 |
|  FSD |  0.086 |  0.044 |  0 |  0.101 |  0.023 |
|  SH |  0.031 |  0.033 |  0.054 |  0 |  0.090 |
|  C |  0.035 |  0.019 |  0.019 |  0.054 |  0 |

**Table 4. Entropy values of lithological states in Karharbari Formation**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  Facies |  E(pre) | E(post) |  R(pre) |  R(post) |  Relationship  |
|  COSD |  1.871  |  1.541 |  0.936 |  0.771 | E(pre) **>**E(post) |
|  MFSD |  1.783 |  1.722 |  0.892 |  0.860 | E(pre) **≥**E(post) |
|  FSD |  1.829 |  1.805 |  0.915 |  0.903 | E(pre) **≥**E(post) |
|  SH |  1.793 |  1.863 |  0.897 |  0.932 | E(pre) **<** E(post) |
|  C |  1.249 |  1.860 |  0.624 |  0.930 | E(pre)**<**E(post) |

E (max) = 2.000 E (system) = 3.97

Facies (COSD): coarse to very coarse grained sandstone, Facies (MFSD): medium to coarse grained sandstone, Facies (FSD): fine grained to ripple laminated fine sandstone, Facies (SH): argillaceous shale and interbedded fine sandstone and shale, Facies (C): coal and shaly coal

E (pre) =Entropy of an individual lithological state in an upward sequence

E (post) =Entropy of an individual lithological state in downward sequence

E (max) =Maximum entropy possible for an individual lithological state

E (system) = Entropy of sedimentation system

R (pre) and R (post) = Relative (normalized) pre- and post entropy of an individual lithological state.

**6.2 Quantitative Results and Geological Discussion**

**6.2.1.Finite Discrete Markov Chain:** Table 1 records the bulk tally count matrix as well as calculated values of three matrices (rij, p ij and d ij) of the Karharbari Formation. Chi-square statistics (ϰ2) also listed, for which the tabulated values at given degree of freedom and at 99.5% confidence limit are high enough to justify the presence of Markov property or in other words cyclicity in the given strata. Markov transition diagram or Facies Relationship Diagram (FDR) in Figure 2 gives only those values of transition probability matrix for which corresponding entries in the difference matrix show positive difference. Highest positive values of d ij matrix link lithologic states distinctly resulting in a strong transition path for lithologic sequence that can be derived as: coarse to very coarse grained sandstone →medium grained sandstone Facies →fine grained ripple laminated sandstone→ argillaceous shale→ coal and shaly coal → coarse to very coarse grained sandstone. This transition path is typical of the coal-bearing sequence and displays a progressive fining of particle size from pebbly, coarse sandstone through fine ripple laminated sandstone to argillaceous shale to coal. A noteworthy feature is a two way transition between medium –fine grained sandstone and coarse grained sandstone, implying that interbedding of the two adjacent lithologies has a greater probability of occurrence in the observed data than would be expected if the lithologies were interbedded randomly. Likewise, a bed of coal seam shows strong preference or memory for the underlying shale and interbedded siltstone and fine sandstone, and the corresponding values of the transition probability are high. Facies Relationship Diagram (FDR) which further indicates that each cycle is asymmetrical beginning with erosional surface and coarse to very coarse sandstone. Individual cycles may vary in thickness from couple of meters to few tens of meters. The finite Markov chain result clearly shows the progressive fining up from pebbly coarse sandstone (basal sandstone) to finer grained sediment is evident This is shown by high probability of passage from coarse to medium grained sandstone indicated by positive (+0.138) d ij values. Binomial probability for this passage of at least *n0bs* success in N trials is 0.080 (Table 2E) which is significant at 10% strengthen the transition passage for facies COSD → MFSD. Invariably, the sandstone is current bedded with variable foreset azimuths. Channel lag deposits at the base grade upward into finer grained sediments invariably ending with either shaly coal or carbonaceous shale/coal intercalations. Cross bedding and lateral channel migration are the rule and indicate a typical channel bar facies. Sediment transport within the channels was mainly in the form of bed load. The upper portion of a typical cycle is dominated by transition probability FSD→ SH (d ij: +0.133). This is suggestive of continuous fining up sequences and supported by typical sedimentary structures, such as ripples, low angle cross bedding, constant foreset azimuths and occasional interrupted by crevasse splay sediments. This may be interpreted probable levees deposition with inter-fingered frequently with back swamp and overbank facies (SH→ C: +0.267) corresponding binomial probability1.746x10-15 (Table 2 E) which is definitely significant at 0.05% level strongly support transition probability from SH to C. The terminal lithological state of such cycles generally consists of shaly coal/coal facies. This possibly reflects in change of depositional sub-environment from levee to back swamp and/ or coal swamp explains this tendency. Interruption of peat formation was common during channel migration and possibly accounted for the patchy occurrence of coal in Karharbari Formation. This coal/shaly coal also exhibit abrupt changes in lithologic transition from C → COSD as indicated by positive d ij value (+0.028). Thus, coal facies overlain by channel sandstone and the resultant sequence are asymmetrical throughout most parts of Karharbari coal measures. The result of Markov chain analysis verifies the former observation made by earlier researchers on the basis of field study that the onset of new cycle is always characterized by an erosional surface which is a rule rather than exception. The coal bearing cycles, charactering fluvial system following Walther’s Law of Facies, represents lateral association of various sub-environments in the similar manner.



**Figure2. Facies Relationship Diagram (FRD) and modal sequence based on statistically significant transition.**

 It is most likely that the Karharbari cycles are auto-cyclic in nature the sediment distributive mechanisms”. The sequence of lithologic transition derived by finite discrete Markov chain in this study is, however, by and large comparable with the cyclic sequences of other late Paleozoic coal measures [30,26,and 28] including the Permian coal measures of lower Gondwana of peninsular India [12,10,and 29] and also Oligocene coal deposits of Northeast India [33].

***6.2.2.* Entropy Analysis*:*** The variation in pre- and post depositional entropy values (Table 4) suggests variable degree of dependency of lithofacies on precursor and influence on successor during Karharbari sedimentation. For pebbly and coarse grained sandstone (facies COSD) E (pre) > E (post), implies with high probability of this lithological state passing up into medium to fine grained sandstone (facies MFSD) but may occur after different lithologic state, as is also recognized in the stratigraphic section. In geological terms, the channel lag or channel bar and inactive channel fill deposits followed each other in upward progression possibly due to anabranching of stream channels. In this process, some truncated cycles were developed as evidence by positive (+) d ij value (Table 2) and FDR (Figure 2). For the lithological states medium to fine sandstone (MFSD) and fine ripple laminated sandstone (FSD) which shows E (pre)≥ E (post) relationship (Table 4) suggest that these lithological facies state following them can be ascertained with more certainty than those preceding them. In simple way these facies exert strong influence by their successor while relatively less influenced by their precursors. In geological term this relationship support geological conclusion that a relatively high probability (Table3) of medium sandstone to fine ripple laminated sandstone may reflect a gradual transition as sediment milieu continued its advance. It is logical to conclude that as the sediment became relatively fine, the corresponding stream channel velocity decreased and ripple laminated sandstone deposits on the adjoining overbank areas. For remaining two lithological sates [i.e. argillaceous shale (SH) and shaly coal and coal (C)] shows E (pre)< E (post) relationship indicate that the dependency of these two facies state on their precursor is much stronger than their influence on their successor. This relationship supports in a statistical way the otherwise geological obvious conclusion that the overbank facies were deposited in the areas abandoned by the stream and peat swamps develop in distant part of the flood plain farther away from the stream channel.



Figure3. (A) Entropy sets derived from Karharbari sequence showing Type A-4 (asymmetrical) cyclic pattern, Circle: possible distribution of entropy sets for asymmetrical cycles after Hattori (1976) (B) Relationship between entropy and depositional environment after Hattori (1976), Dot: Entropy for Karharbari sequence, Talchir sub-basin.

The plot of the E (pre) and E (post) values for each lithological state is given in the Figure 3A of Hattori’s diagrams, which most closely (though not exactly) follows the expected for an asymmetrical cyclic sequence (Type A-4 category) of [25]. Among entropy sets those for facies COSD, MFSD, FSD and SH (pebbly to coarse sandstone, medium to fine grained sandstone, fine sandstone to ripple laminated fine sandstone and argillaceous shale) are located along the diagonal line; there is a fair probability that these lithological states tend to occur as a symmetrical cycle, that is, alternation in the succession as observed in outcrop sections. The entropy set with respect to facies C (shaly coal and coal) deviates from the general distribution (figure 3) is possibly due to small values of pre- and post entropy. This departure in the entropy set for shaly coal and coal (facies C) maybe explained in terms of the opinion expressed by [34] that coal forming environment whether back swamp or marsh is not a normal feature of alluvial flood plains and may develop either locally in a part of the flood plain or occupy the entire basin overlying the deposits of various sub-environments and probably neither was this during the Karharbari sedimentation. This may possibly explain why in Karharbari sedimentation carbonaceous / coal seam (facies C) do not occur regularly with each alluvial-fluvial cycle.

The value of E(system) statistics for the five lithological states data was 3.979 , compared to maximal and minimal values of 4.322 and 2.321.When the E(system) value for the Markovian sequences are evaluated in terms of broad depositional environment by plotting E(system) against the number of lithological states, the point occurs within the field allotted to fluvial-alluvial of [25] Figure 3B thus corroborating the views expressed by earlier workers that the Karharbari sedimentation took place by lateral shifting braid channels in response to varying discharge and rate of deposition caused the development of asymmetrical (Type A- 4) cycles.

To sum up the finite Markov chain and entropy studies strongly support that the sediments of the Karharbari succession was deposited by Markovian mechanism and as a whole represents alluvial-fluvial sedimentation deposited in a predictable cyclic arrangement of lithofacies. The statistical results concur with observed sedimentological evidence of depositional environment. Fining upward cycles correspond to the development of basal channel bars, changing into levees, then into back swamps which are topped by coal swamps. Frequent interbedding of coal and argillaceous/ carbonaceous shale in the upper part of the coal measures calls for periodic flooding and flushing of fine clastics into coal forming swamps causing interruption of peat formation. Cyclical deposition is explained by wandering channels in response to varying discharge and rate of deposition, their frequent lateral shift of channel course with consequent overstepping of adjacent sub-environments of the flood plain, a common feature rather than exception in river basins, may favorable explain for the formation of symmetrical cycles, with a paucity of fine clastics, characterize the Karharbari succession in this alluvial-fluvial depositional model.

 **7. Quasi-Independence Analysis**

The finite Markov model, indeed, does not correctly verify the problem of randomness tests as it ignores the lack of lithofacies transitions into themselves (the presence of zero values at the diagonal of the tally count matrix). Subsequent some authors [7, 17, and 18] proposed the improved versions, where the quasi independence model of [17], a statistically rigorous method for evaluating incomplete matrices- abbreviated as Power and Easterling (PE). The PE model can be used to generate the expected cell values for the independent trials matrix.

**7.1 Supplication of the PE model:** The modified and improved Markov chain process after Power and Easterling (PE) used in this study incorporates following successive steps.

Structuring of the number of observed transitions matrix (f *ij*) where *i, j* corresponds to row and column letters refers to lower and upper bed respectively; (ii) matrix of the number of expected transitions in a random sequence (E *ij*) derived by using an iterative procedure; (iii) matrix of differences between the number of observed transitions and the number of expected transitions in a random sequence (d *ij*) = (f*ij*) - (E *ij*). This provides a framework for identifying large difference between observed (f *ij*) and expected transition frequencies (E *ij*). The hypotheses were tested for consistency with random sequence and for randomness of specific transitions between lithofacies (significance test of specific elements of the matrix (d *ij*). Among the latest improvements of Gingerich and Read (GR) method the most credible is the Power and Easterling’s [1982] proposal for more discussion readers should see in [35] here abbreviated as PE; (iv) Using the values of f *ij* and E *ij* in the expression for ϰ2 yields a statistics which is approximately distributed as a Chi-square variable with *(n-1)2-n* degrees of freedom. The larger theϰ2 value for a given value of *n*, the stronger the evidence in favor of the Markovian model of lithologic transition, i.e. for the presence of cyclicity; (v) comparison of observed f *ij* and expected E *ij* number of transitions; (vi) Whether or not a difference represents a ‘significant’ departure from quasi-independence depends on the size of the probabilities being estimated and on the amount of data involved in the estimates. To alleviate this problem, [36] introduced normalized difference matrix (Nd ij), which can be calculated as Ndij = (f ij-E ij)/ √Eij where large differences (+ values) between observed and expected transition frequencies and referred as ‘dominant’ transition. The significant excess of transitions is conventionally displayed in a diagram, with arrows connecting the lithological states for which there are positive differences and regarded as ‘ideal’ or fully developed cycles, which facilitates the geological interpretation. The inferred assumption underlying this procedure is that the expected matrix represent ‘noise’ subtracting this from the observed matrix one can filter the observations for randomness or noise. ij

The calculation method ofϰ2 statistics used for the randomness test of a sequence is identical in the GR and the PE methods, as follows

 n n

 ϰ2 = $ \sum\_{}^{} $∑ = (f ij – E ij) 2/ E ij

 i=1 j=1

In the PE method the number of degrees of freedom is defined as *(n-1)2-n* where; *n* = number of lithofacies states, d ij = elements of difference matrix, E ij = elements of expected transition matrix. In the GR method the number of degrees of freedom is defined as *n2-2n.*

In the Power and Easterling (PE ) method, E ij be the expected value of the number of transitions from state ‘*i’* to state *‘j’* in quasi-independence is calculated as:

 E ij = ai bj, for i ≠ j

 = 0 for *i =j*

where *ai* and *bj*, denotes frequency of individual in the ith and jth row, respectively and the values of *ai*and *bj*  are determined with relevant iterations.

Let *m* represent the number of rows/columns in the transition frequency matrix, *ni+* be the sum of *i* row, *n+j* be the sum of *j* column and ω be the required accuracy of iterations. Thus first iteration:

 *ai (I) =*  *ni+* / (m-1),  *i*= 1,2,3,….*m*

 bj(I) = n+j  / ∑ ai , j =1,2,3,…m

  *i≠j*

Similarly Ith Iteration:

 ai (I) =  ni+ / ∑ bj (I-1)i=1,2,3,…m

  *j≠i*

 bj(I) = n+j  / ∑ a i(I) j =1,2,3,…m

  *i≠j*

where *ni+* and *n+j* are the row *i* and column *j* totals, respectively.

Iteration may be run until the following condition is obtained:

 *ai (I)* – *ai (I-1)* ≤ ω for *i*= 1,2,3,….*m*

 bj(I) \_ bj (I-1)≤ ωfor j =1,2,3,…m

**7.1.2. Implementation of Quasi-independence Model:** An application of quasi-independence model is presented of estimation method of the expected number of facies transitions. True data originate from the sector III of the Barakar Formation of Lower Gondwana succession, which represents the fluvial series of the Early Permian of Bellampalli coal basin, Pranhita Godavari Valley Basin. The succession includes 190 events of four lithological facies: coarse to medium sandstone (COSD – 46 events), interbedded fine grained sandstone and shale (FSD – 14 events), carbonaceous shale and argillaceous shale (SH – 67 events) and shaly coal and coal (C – 63 events). The observed transition count matrix f ij between these lithofacies along with matrices of the expected numbers of transition in a quasi-independence calculated by PE method (E ij), normalized difference matrix (N dij) and chi-square (ϰ2) are shown in Table 5.

Table5. Bulk tally count and calculated values of expected cell value and normalized difference matrix and chi-square matrices of Barakar Formation, Bellampalli coal basin by PE method.

(A). Transition count matrix ( *f ij* )

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  facies |  COSD |  FSD |  SH |  C |  *n+i* |
|  COSD |  00 |  08 |  23 |  15 |  46 |
|  FSD |  06 |  00 |  06 |  02 |  14 |
|  SH |  18 |  03 |  00 |  46 |  67 |
|  C |  22 |  03 |  38 |  00 |  63 |
|  *n+j* |  46 |  14 |  67 |  63 |  *n***++=**190 |

 (B). Expected cell value matrix (Eij)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  facies |  COSD |  FSD |  SH |  C |  *n+i* |
|  COSD |  00 |  3.063 |  22.984 |  19.872 |  45.919 |
|  FSD |  3.064 |  00 |  5.858 |  5.072 |  13.994 |
|  SH |  22.925 |  5.858 |  00 |  37.978 |  66.753 |
|  C |  19.861 |  5.072 |  30.059 |  00 |  63 |
|  *n+j* |  45.850 |  13.985 |  66.981 |  62.922 |  *n***++=**189.258 |

(C) Normalized difference matrix (Ndij)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  facies |  COSD |  FSD |  SH |  C |
|  COSD |  00 |  **+2.820** |  **+0.002** |  -1.093 |
|  FSD |  **+1.678** |  00 |  **+0.062** |  -1.362 |
|  SH |  -1.038 |  -1.813 |  00 |  **+1.305** |
|  C |  **+0.479** |  -0.923 |  **+1.436** |  00 |

 (D) Chi-square matrix (ϰ2)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  facies |  COSD |  FSD |  SH |  C |
|  COSD |  00 |  6.714 |  00 |  1.194 |
|  FSD |  1.436 |  00 |  0.003 |  1.862 |
|  SH |  1.065 |  1.393 |  00 |  1.694 |
|  C |  0.208 |  1.112 |  1.659 |  00 |

 ϰ2 Degree of freedom Limiting value at 0.05% significance level

 18.326 5 11.070

**7.1.3. Quantitative Results and Geological Discussion**

***Quantitative Results*:** Table5 records the bulk transition tally matrix as well as calculated values of expected cell value (Eij), normalized difference (Ndij) and chi-square (ϰ2) matrices for the Barakar of the Bellampalli coal basin. Indeed, there is a strong tendency of Markovian property or in other words cyclicity in the Barakar succession at a given degree of freedom and 0.05 % confidence level. Figure 5 show Markov transition diagram based on positive values of normalized difference matrix. Highest positive values of Ndij matrix link lithologic states distinctly resulting in a strong transition path for lithologic sequence that can be derived as: coarse to medium sandstone (facies COSD), interbedded fine sandstone and shale (facies FSD), argillaceous and carbonaceous shale (facies SH), shaly coal and coal (facies C), carbonaceous shale and argillaceous shale (facies SH) and coarse to medium sandstone (COSD) resulting asymmetrical cyclothems. Furthermore, the diagram represents a graphic visualization of matrices of differences between frequencies of expected and observed transitions enables the reconstruction of model sequence, i.e. such which shows a genetic implications resulting from the nature of depositional mechanism. Individual model cycles may vary in thickness from a couple of meters to few tens of meters. This transition path is typical of the coal-bearing and by and large display a progressive fining of particle size from coarse sandstone through fine sandstone to shale, then coal. A noteworthy feature with this sequence is a two way transition between coarse and fine sandstone, implying that interbedding of the two lithologies has a greater probability of occurrence in the observed data than would be expected if the lithologies were interbedded randomly. Likewise, a bed of shaly and coal shows a strong preference or memory for the underlying shale and the corresponding positive values of the normalized difference matrix. It was concluded that transitions of depositional intervals and lithofacies are not independent but exhibit a first-order finite Markov property. Evidently, the quantitative results imply recurrence of the corresponding depositional environments both across the basin and through time,

***7.1.4.*****Geological Discussion:** Quantitative results suggest that a complete Barakar cycle usually show a progressive fining up from coarse, basal sandstone to finer grained sandstone is evident. This is shown by high probability to passage from coarse (COSD) to fine sandstone (FSD) indicated by high positive Ndij value (COSD→ FSD = 2.820; FSD → SH =0.62; SH → C = 1.305; see Figure 4 and Table 5). It is dominant lithofacies, which constitutes about 65% by volume of the coal bearing Barakar Formation. Invariably, the sandstone occurs as channel to sheet like and multistory bodies showing erosional base and flat-top is current bedded both planar and trough cosets. Cross bedding and lateral channel migration or aggradations are the rule and indicate of a typical point bar facies. This set up is suggestive of channel shoaling of a typical point bar facies [38, 4, and 6]. Multistory sandstone bodies in alluvial-fluvial



**Figure 4. Method of the construction of model and modal sequence, Barakar Formation, Bellampalli coal basin**.

environments are probably formed when the rate of migration within aggrading channel belt is large enough to cause superposition of channel bars, before the channel belt is abandoned [39] whereas sheet like sandstone bodies are formed when subsidence rates is slow than sedimentation rate. Sediment transport within the channels was mainly in the form of bed load as outlined by [40] on the basis of paleohydrological studies. The coarse and fine grain grade members (facies: COSD & FSD) are deposits which in texture and sedimentary structures compare very closely with the channel sediments of modern rivers [41, 42, and 43]. The upper portion of model cycle is dominated by transition probability FSD → SH (Ndij: =0.062).This is suggestive of continuous fining up sequences and supported by typical sedimentary structures, such as ripples, low angle cross bedding and occasionally interruption by crevasse splay sediments. This may be interpreted as deposition by vertical/lateral accretion on top of channel bars during a low water stage or as levee deposition which inter-fingered frequently with back-swamp facies. Following this phase of sedimentation, sediment supply was curtailed; lacustrine conditions of stagnant water may have developed in then low lying areas beyond channel and overbank sub-environments where swamp or marshy conditions began to grow across the plain. Similar peat forming environments may also have developed in areas of abandoned channels. Indeed as expected, argillaceous and carbonaceous shale show a greater probability to pass upwards into shaly coal and coal (Ndij =1.305) than sandstone or shale. The occurrence of coal seams at the top of the fining upward cycles, therefore, suggests cessation of alluvial-fluvial environment and conversion of the flood plain into lake or swamp. Furthermore, upward linkage of coal and shaly coal facies which represents the top unit of the Barakar cyclical units has more preference for underlying argillaceous and carbonaceous shale (facies SH), resulting in the asymmetrical fining cycles, implies a gradual encroachment of coal swamp by adjacent back swamp and levees sub-environments as a consequence of slow and gradual lateral shift of channel course across the alluvial plains. These episodes of undisturbed peat accumulation over a considerable time span was repeated several times (6 coal seams of workable thickness, ranging in thickness from 5 to 15m, occurs and accounts about 5% of the Barakar succession. In addition, there are several impersistent coal beds of less than a meter in thickness which probably represent their formation in inter-channel and/ or distal flood plains or frequent shifting of channels may have prevented development of thick peat swamps. Initiations of second and subsequent fining upward cycles in study area commonly start with the deposition of coarse-medium sandstone (facies COSD). This may be attributed to random change in the depositional mechanism [4] manifested by the development of some kind of channel system after phases of peat accumulation.

As the Barakar cyclothems are non-marine and fluvial in nature, the cyclothems might have been formed by sedimentary control theory in the form of sedimentary distributive mechanism of lateral migration of streams across the alluvial plain [44, 9, and, 10] or drainage diversion [45] caused in response to intrabasinal differential subsidence or repeated slow and rapid subsidence of the depositional basin as suggested recently [46].

**8. Markov Reversibility and Marginal Homogeneity**

Complex model for contingency tables have received an increasing interest in recent years from researchers especially in biological, physiology and geological sciences play important roles, as they highlight relationships between categorical variables [47]. Contingency tables with the same row and column classification are often called square contingency table and arise in many fields such as social mobility data, biomedical research and lithological data in geology and so on. Typically, the analysis of square contingency tables considers the symmetry issue rather than independence because observations tend to concentrate on or near the main diagonal. Therefore, many statistician and geologists have considered various symmetry (S), quasi-symmetry (QS) and marginal homogeneity (MH) models, and symmetry is one of the important topics for the analysis of square contingency tables and there are number of well described examples of tables that have modeled as quasi-symmetry [48,49,50, and51].The model of quasi-symmetry (QS) requires that the expected marginal total of any row of the square table not the same as the expected marginal total for the corresponding column. [49] provided a sample test (chi-square statistics) for marginal homogeneity (MH) which indicates the equivalence of marginal probabilities.[52] developed quasi-symmetry (QS) based on without requiring knowledge of the limiting probabilities which has since become widely accepted for the analysis of square contingency tables are illustrated in the geological study. The quasi-symmetry (QS) model bridges symmetry (S) and marginal homogeneity (MH) in square tables. If the symmetry model holds, then the marginal homogeneity model holds. However, the inverse is not always true. The quasi-symmetry model can be used to show that the symmetry (S) model holds if and only if both (QS) and (MH) models hold than the underlying process is a reversible. Thus acceptance of a simple Pearson chi-square test for symmetry (S) model of the tally matrix is sufficient to show the absence of Markov cyclicity in a sedimentary succession

An important feature of the finite Markov chain is the property of reversibility generally found in cyclical lithologic transitions in stratigraphic succession [59, 50]. A direct relationship between reversible Markov chain model and quasi-symmetry (QS) model has been previously noted; this relation is not easily understood until quasi-symmetry (QS) model is considered as an unequivocal mathematical property. In geological words, a Markov chain sequence is reversible if and only if the tally count matrix of transition counts possess quasi-symmetry (QS) or satisfies balance equation of Kolmogorov criterion.

In this paper we illustrate the use of the quasi-symmetry (QS), symmetry (S) and marginal homogeneity (MH) models in the analysis of square categorical data that summarized the constituents lithologic of cyclothems. Furthermore, in order to understand the nature of sedimentary successions and in particular to identify the presence or absence of cyclical sequences, quite a number of sedimentary successions have been counted to determine their Markov property. In reviewing classical examples of different geological ages around the world of these published tabulations of sedimentary successions, it quickly become clear that most of these tabulations have marginal homogeneity (MH) and the square tables are either symmetric (S) or non-symmetrical and the sequence confirms to a reversible or non reversible Markov process.

**8.1****Marginal Homogeneity***.* The marginal homogeneity model (MH) is defined by

 *fi+ = f+i* (i =1,2,3……..r)

where *fi+*= ∑j=1 *fij* (*i.e.* row sums ) and *f+*i = ∑j=1 *fji* (*i.e.* column sums). The MH model indicates that the marginal distribution of the row variable is identical to the marginal distribution of the column variable. In fact MH model must hold provided that the succession of sedimentary states forms a continuum in which each state is assessable from at least one other state. This is commonly encountered condition because sedimentary sections encountered during detailed stratigraphic traverses or from long continuous lithological profiles.

If there is any doubt, a chi-square test (ϰ 2) can be used to check for identity between row and column sums for the Continuous Time Markov chain (CTMC), that is:

 ϰ 2= ∑ (*fi*+─ *f*+*i*) 2 / ( *fi*++*f*+*i* ─2 *fii*) (1)

This statistics test asymptotically distributed, under the assumption of marginal independence, as (ϰ 2) with *n-1* degrees of freedom. Acceptance of the hull hypothesis H0: *fi+ = f+*i (*i.e.* the two samples have the same marginal distribution) at least 5% confidence level. If only transitions between sates have been counted, then the diagonal elements of the tally matrix are either absent or zero *i.e*. Discrete Time Markov chain (DTMC) so that the above χ2 test reduces to:

 ϰ 2= ∑+ (*fi*+─ *fi*)2 / (*fi*++ *f*+*i*) (2)

According to [52, 53] the above chi square test that it is only an approximation but it should be adequate for most instances encounter in study of cyclic successions [49, 50].

***8.1.2* Quasi-Symmetry**.Caussinus [54] developed quasi-symmetry model (QS) which bridges symmetry (S) and marginal homogeneity (MH) models in square contingency tables.The quasi-symmetry (QS) model defined as the probability that an observation falls in *i* and *j* cell.The QS model requires that the expected marginal total of any row of the matrix not the same as the expected marginal total for the corresponding column. The QS model can be used to show that the S model holds if and only if both the QS and MH model hold, the inverse is not always true.

Consider an *n* x *n* square contingency table with the same row and column classifications.Let X and Y denotes the row and column variables, respectively, and let *fij denotes* the probability that an observation will fall in the cell (*i, j*) of the square table. The symmetry (S) model which is defined as  *f ij = fji* for *i* =1, 2, 3….R; *j* =1, 2, 3…. R indicates symmetry structure of cell probabilities. Additionally MH model indicates the equivalence of marginal probabilities and is defined as *fi*+= *f*+*i for i*=1 ,2, 3….R where *fi+*= ∑j=1 *fij* and *f+*i= ∑j=1 *fji* [52]. The quasi-symmetry (QS) model is defined as

  *fij fjk fki= fik fkj fji* for any *i, j , k* with 1*≤ i, j, k ≤ n* (3)

The above inequality can be simply stated as a Markov sequence is reversible if and only if the tally matrix transition counts possess quasi-symmetry and conversely Markov sequence is irreversible.

In the case where the transition counts matrix (f*ij* ) possess marginal homogeneity and the entries in the square contingency table are subject to random error, a Pearson chi-square test is available for testing for the complete symmetry model (S) *pij= pji*, *i ≠ j* is tested by calculating following test statistics as applied by [48, 49 and,50]

 ϰ 2 = ∑ ∑ *(fij – fji)2 / (fij + fji)* (4)

 *i*>j

This is asymptotically distributed as χ2 with n (n-1) /2 degrees of freedom. Following the convention of Fisher, acceptance of the null hypothesis will be assumed at the 5% confidence level and rejection of the null hypothesis at the 1% confidence level.

**8.1.3. Geological Application.**From among a large number of published Finite Markov chain analyses (DTMC) the following table of 2378 transition counts from 32 borehole logs were counted from the Lower Permian coal-measures of the Tanzanian Mchuchuma basin [28] were selected to demonstrate the application of marginal homogeneity (MH) model, quasi-symmetry (QS) and symmetry (S) model in geological data.

Tally Count Matrix of Tanzanian Mchuchuma basin (*f ij*)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| facies | A | B | C | D | E | E+F | fi+ |
| A | 0 | 435 | 120 | 46 | 22 | 06 | 629 |
| B | 315 | 0 | 284 | 86 | 25 | 08 | 718 |
| C | 26 | 155 | 0 | 134 | 16 | 37 | 468 |
| D | 48 | 41 | 34 | 0 | 36 | 98 | 257 |
| E | 32 | 19 | 08 | 06 | 0 | 41 | 106 |
| E+F | 99 | 56 | 17 | 04 | 24 | 0 | 200 |
| f+i | 520 | 706 | 463 | 276 | 123 | 190 | *n*++= 2378 |

A=coarse sandstone, B= medium sandstone, C. fine sandstone, D. shale, E. mudstone and E+F= coal=mudstone

The visual inspection of the entries in the square contingency table shows that the transition count matrix has lack of perfect marginal homogeneity indicating the lithofacies states profiles have not properly counted and this is not confirmed by tested for marginal homogeneity (equation 2), the result gives χ2 =12.667 where χ2 (95%) = 11.09 for 5 degrees of freedom but not statistically significant. Presence of perfect marginal homogeneity is therefore statistically not established and the sedimentary section has not been properly counted. If the square tally count table is then tested for symmetry (S) by chi-square test (equation 4), the quantitative result gives χ2 = 1407.041 where χ2 (1%) =37.70 and degree of freedom=15 and is on the basis of chi-square result it is a case of quasi-symmetry (QS). Consequently, the Lower Permian coal-measures of the Tanzanian Mchuchuma basin conforms to a non-reversible Markov process, hence the coal-measures succession possess Markovian mechanism, and that the sequence represents cyclic sedimentation corroborating inferences deduced by earlier workers on the basis of subjective approach.

Next consider a coal-bearing Barakar sedimentary sequence of Saharjuri sub-basin of Koel-Damodar Valley Gondwana master basin of peninsular India, normally consists of five lithologies, were condensed into three different lithologic states as follows: sandstone (A), shale (B) and coal (C) [19]. The condensation of lithologies was necessary to avoid risk of error which is not unlikely if lithologies are subdivided minutely and arbitrary as in the given borehole logs. A traverse along such a sedimentary sequence of 102 counts will then yield the following transition count matrix following Continuous time Markov chain (CTMC).

Transitional Count Matrix Barakar Formation, Saharjuri (*f ij* )

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lithofacies | sandstone | shale | Coal  | fi+ |
| sandstone | 19 | 15 | 04 | 38 |
| shale | 11 | 13 | 09 | 33 |
| coal | 07 | 08 | 16 | 31 |
| f+i | 37 | 36 | 29 | *n*++ =102 |

Simple inspection of the row and column sums is sufficient to show nearly perfect marginal homogeneity suggesting a properly counted sequence. A scrutiny of the non-diagonal entries also shows that the square table is symmetrical and formal testing by equation (3) confirms the result with χ2 =2.571 where χ2 =5.99 (5%) is with 2 degree of freedom. Consequently the coal bearing Barakar sequence conforms to a reversible Markov process indicating that the reverse and forward sequences are same.

Now consider the cyclothems presented [23] for a thick Paleocene section of alluvial sediments forming Polecat Beach Formation in Bighorn basin, Wyoming. The reported Paleocene alluvial sequence can be treated as a simplified four unit’s cyclothems. The units are A. sandstone, B. mudstone, C. limestone and D. lignite. The observed square tallies are as follows:

Transition Count Matrix of Polecat Beach Formation

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Litho-units | A | B | C | D | fi+ |
| A | 0 | 37 | 03 | 02 | 42 |
| B | 21 | 0 | 41 | 14 | 76 |
| C | 20 | 25 | 0 | 00 | 45 |
| D | 01 | 14 | 01 | 0 | 16 |
| f+i | 42 | 76 | 45 | 16 | *n*++ =179 |

Observe first of all that the square tally table has perfect marginal homogeneity (MH) indicating the cyclothymiacs units have been properly counted. So a test for marginal homogeneity is not required. Furthermore inspection indicates the table in non-symmetrical and formal testing confirm the result with χ2 =22.19 where χ2 (5%) =12.59, (1%) =16.81 and 6 degree of freedom. Thus this Paleocene sequence cannot be a reversible Markovian sequence; hence the statistical result supports the geological interpretation that Polecat Bighorn Formation possesses cyclicity.

**9. Markov Reversibility**

Markov chains are a versatile probabilistic tool for modeling stochastic processes that have achieved widespread application in the quantitative sciences. A Markov sequence is a series of states generated by mathematical process in which transitions from one state to another state within a finite number of possible events. It is collection of different states and probabilities of a variable, where its future state is substantially dependent on its immediate previous state. In the case of a Markov reversibility process, a tally matrix of the sequence of states in the forward direction is indistinguishable from the tally matrix counted in the backward direction.

***Reversible Markov Chain*:** A Markov process is called a reversible markov chain or process if it satisfies the detailed balance equations. These equations require that the transition probability matrix *p* for the Markov process possess a stationary distribution π such that

 *πi pij* = *π j p* ji (Balance equation)

where p*ij* is the Markov transition probability from state *i* to state *j* , *i.e.* pij = p(xt = j │ x *t-1* = *i* ), and π *i* and π *j*  are the stationary probabilities of being in state *i*  and *j*, respectively. When pr (xt-1 = *i)* = πi for all *i ,* this is equivalent to the joint probability matrix, pr (xt-1 = *i,│ xt = j*) being symmetric in *i* and *j*, or symmetric in *t-1* and *t*.

The above discussion can be summarized in the form of a theorem.

**Theorem 1**: A stationary process if and only if there exist a positive collection of number (π i) to unity such that πi pij = π j p ji for all i and j, whenever such collection does not occurs, the Markov process is reversible.

The above detailed balance equation means that the probability of seeing a transition from lithological state *i* to lithological state *j* equals as seeing a transition from lithological state *j* to *i* lithological state. In the case *i* = *j* the two sides of the balance equation are identical and hence carry no geological information. Thus, the detailed balance equation allows us to determine if a finite Markov process is reversible (non cyclical sequence) or irreversible (cyclic sequence) based on the transition probability matrix and the stationary probabilities.

**Theorem 2**. A finite Markov chain with transition matrix P is reversible if π x P is symmetrical, where x means component-wise multiplication.

Proof. Let R = π x P

 π1 π2 π3

 *π* = π1 π2 π3

 π1 π2 π3

 *p*11 *p*12 *p*13

 *\** *P = p*21 *p*22 *p*23   1 *p*31 *p*32 *p*33

π1*p12 π2p12 π3p13*

 *π \* P* = *π2p11 π2p12 π2p23*

 *π3p11 π3p12 π3p23*

 If we get *π \* P* a symmetric matrix, i.e., *πi* pij = *π* j pji for all *i, j,* then the detailed balance equation

is satisfied hence the transition matrix P is reversible (cyclic), otherwise non-reversible (non-cyclic)

 ***9.1 Geological Application****:* To illustrate the use of detailed balance equation in geological studies, an example is taken from the Barakar coal measures of Bellampalli coal basin, Pranhita-Godavari Valley Gondwana Basin, Andhra Pradesh. The Barakar Formation comprising a spectrum of sedimentary facies including gritty sandstone, sandstone, siltstone, argillaceous and carbonaceous shale and coal. The following table of transition count matrix and transition probability matrix obtained by DTMC method of aforesaid sedimentary succession [55].

 (A).Transition count matrix of Barakar Formation (*f ij*)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| lithofacies | A | B | C | D |
| A | 0 | 08 | 113 | 35 |
| B | 08 | 0 | 03 | 01 |
| C | 78 | 03 | 0 | 65 |
| D | 59 | 01 | 79 | 0 |

(A). Coarse-medium sandstone; (B) Interbedded sequence of argillaceous shale and fine sandstone; (C) carbonaceous shale, and (D) coal.

 (B).Transition probability matrix (pij)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| lithofacies | A | B | C | D |
| A | 0 | 0.051 | 0.724 | 0.224 |
| B | 0.666 | 0 | 0.250 | 0.083 |
| C | 0.534 | 0.020 | 0 | 0.445 |
| D | 0.424 | 0.007 | 0.568 | 0 |

The stationary distribution vector of transition probability matrix (pij) is *π =* (0.332, 0.016, 0.400, and 0.252). Now component wise calculation using relation R = *π \* P* we get following reversibility matrix.

(C).Reversibility matrix of Barakar Formation (R)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| lithofacies | A | B | C | D |
| A | 0 | 0.016 | 0.240 | 0.074 |
| B | 0.012 | 0 | 0.004 | 0.001 |
| C | 0.213 | 0.008 | 0 | 0.178 |
| D | 0.106 | 0.001 | 0.142 | 0 |

Simple inspection of rows and column of the reversibility matrix shows that table lack symmetry. By theorem 1 if the reversible transition matrix is non symmetrical then the succession possesses Markov cyclicity. The table is non symmetrical and formal testing by chi square test (equation 3) as applied by [47] confirm the result with (ϰ 2) =88.112 where ϰ 2(5%) = 12.59, ϰ 2(1%) =22.45 with degree of freedom =6. Thus the statistical result support above contention that the geological interpretation that the Barakar succession is a cyclothems *i.e.* deposited as Markovian clock.

**Theorem3. (Kolmogorov Criterion).** An n x n Markov chain is reversible if and only if the product of one-step transition probabilities along any finite closed path of length more than two is the same as the product of one-step transition probabilities along the reversed path. In other words

  *pij pjl plk pki = pik pkl plj pji*

For all finite sequences of state

This is the definition of reversibility introduced by Andrei Kolmogorov and known as Kolmogorov criterion as a necessary and sufficient condition for Markov chain or discrete time Markov chain (DTMC) to be reversible directly from its transition probability matrix. The criterion requires that the products of probabilities around the loop *i* to *j* to *l* to *k* returning to *i* must be equal.

***9.2. Geological Application****:* A practical geological application of the Kolmogorov criterion can be illustrated by the following table of transition counts for alluvial sediments of Beaufort Group of the Karoo basin, South Africa [57].

 Transition count matrix of Beaufort Group (*f ij*)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E |
| A | 0 | 39 | 05 | 21 | 03 |
| B | 26 | 0 | 28 | 27 | 01 |
| C | 01 | 30 | 0 | 06 | 04 |
| D | 10 | 26 | 16 | 0 | 06 |
| E | 01 | 04 | 00 | 03 | 0 |

A= Mudstone clast conglomerate, B= Structure-less sandstone, C= Plane bed sandstone, D= Cross-bedded sandstone, E= Mudstone

Transition probability matrix of Beaufort Group (pij)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E |
| A | 0 | 0.573 | 0.073 | 0.309 | 0.044 |
| B | 0.317 | 0 | 0.341 | 0.329 | 0.012 |
| C | 0.024 | 0.731 | 0 | 0.146 | 0.097 |
| D | 0.172 | 0.448 | 0.276 | 0 | 0.103 |
| E | 0.125 | 0.500 | 00 | 0.375 | 0 |

 Substituting the transition probability matrix (pij) values of in Kolmogorov criterion of above inequality,

  *pij pjl plk pki = pik pkl plj pji*

we have

 0.573x0.341x0.146x0.103x0.125 = 0.375x0.276x0.731x0.317x0.044

From A 5x5 transition probability matrix, it turnout that only one detailed balance equation has been checked and found to be equal suggesting that the probability matrix is *reversible* which in turn supports’ the geological interpretation of cyclical deposition of Beaufort Group within the Banksgaten Sandstone of Karoo Group as elaborated discussed by [57].

 **Theorem 4.** A stationary Markov chain is reversible if and only if the matrix of transition probabilities can be written as the product of a symmetric and a diagonal matrix i.e., P =SD where S is a symmetric matrix and D is a diagonal matrix such that ∑j pij =1

Proof:

 π1 0 0 p11 p12 p13

Let D = 0 π*2* 0 and P = *p21 p22 p23*

 0 0 π3 p31 p32 p33

Then according to above theorem S = P D-1

 p11 p12  p13 1/ π1 0 0

 = p21 p22 p23 χ 0 1/ π2 0

 p31 p32 p33 0 0 1/ π3

 p11/ π1 p12/ π2 p13/ π3

 p21/ π1 p22/ π2 p23/ π3

 p31/ π1 p32 / π2  p33/ π3

As we have shown above that if *πi j pi* j *=* *πj pj i*, (balance equation) then the Markov chain is reversible. Thus we get

 p11/π1 p21/π2 p31/π3

 S = *p21*/π1 *p22/*π2 *p32/*π2

 p31/π1 p32/π2 p33/π3

This is clearly a symmetric matrix. Since S =PD-1, we get P = S D as required.

***9.3. Geological Application****:* To illustrate the use of Theorem 4 consider the late orogenic middle molasses Siwaliksequence which consists of a spectrum of fluvial dominated depositional unitsof COSD (coarse sandstone), MFSD (medium to fine sandstone), SLSD (interbedded sequence of siltstone and fine sandstone) and MDST (mudstone). Given as, the following observed transition count and transition probability matrices [58],

Transition count matrix (*f ij*)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | COSD | FMSD | SLSD | MDST | *fi+* |
| COSD | 0 | 28 | 05 | 02 | 35 |
| FMSD | 17 | 0 | 23 | 05 | 45 |
| SLSD | 10 | 07 | 0 | 14 | 31 |
| MDST | 07 | 10 | 04 | 0 | 21 |
| *f+i* | 34 | 45 | 32 | 21 | *n*++ = 132 |

Note that the observed transition matrix has perfect marginal homogeneity indicating a properly counted sequence. Simple inspection suggests that the observed table lacks symmetry and this is confirmed by testing with chi-square equation (3), the result gives χ2 =22.19 where χ2 (15) = 16.81 with 6 degrees of freedom.

Transition probability matrix (pij)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| lithofacies | COSD | FMSD | SLSD | MDST |
| COSD | 0 | 0.800 | 0.142 | 0.057 |
| FMSD | 0.377 | 0 | 0.511 | 0.111 |
| SLSD | 0.322 | 0.225 | 0 | 0.452 |
| MDST | 0.333 | 0.476 | 0.190 | 0 |

Now calculate out its stationary probability (π) from probability matrix (pij). The values are

 (π) = (0.346, 0.318, 0.197, and 0.138)

Substituting the values in equation S = PD-1, we have

0 0.800 0.142 0.057 1/0.346 0 0 0

0.377 0 0.511 0.111 χ 0 1/0.318 0 0

0.322 0.225 0 0.452 0 0 1/0.197 0.333 0.476 0.190 0 0 0 0 1/0.138

 0 2.512 0.721 0.413

 S = 1.894 0 2.588 0.804

 0.934 0.707 0 3.275

 0.963 1.497 0.964 0

The above result shows that S matrix is a non-symmetrical hence P is a non reversible matrix suggesting that the Siwalik molasses succession of the study area possess Markov cyclicity hence supporting unreservedly Khan’s [58] geological interpretation based on the DTMC and entropy analyses.

**10. CONCLUSIONS**

Finite Markov chain stochastic process together with binomial functions has been used to objectively distill the actual lithofacies transition trend which was masked by several erosion truncation surfaces in the coal bearing Karharbari and Barakar sandstone. The transition pattern was found to be Markovian clock, indicating a link between the underlying and overlying lithofacies, in turn, and the environments of deposition. The fining upwards cycle shown by the logged boreholes data indicated deposition in alluvial-fluvial braided river systems. Hattori’s entropy functions indicate that cycles are mostly asymmetrical in nature and developed in alluvial- fluvial environments. Fining upward cycles correspond to the development of basal channel bars, changing into levees, then into back swamps which are topped by coal swamps are of auto-cyclic in nature. The statistical results concur with observed sedimentological evidence of depositional environment. Finite Markov chain stochastic process and Entropy functions have been shown in this study to be an appropriate quantitative methods for defining lithofacies trend with precision The asymmetric cyclic sequence with coal as terminal facies, lateral pinching and splitting of coal seam are features which suggest allocthonous origin Permian Gondwana coal. The geological examples presented above demonstrate that all statistical methods produce almost identical results; hence, all provide the same conclusions after the randomness test calculated. It is suggested, therefore that the methods may be applied in oil well lithofacies data analysis to determine and define lithofacies cycles or trends. When such cycles are established in a well they can be further used for local and field-wise correlation. They would also be useful to predict stratigraphy in unexplored geologically areas in a basin.

Contingency tables highlight relationships between categorical variables. Typically, the symmetry or marginal homogeneity of a square table is evaluated. If transition counts are made and tabulated among different discrete and or continuous lithofacies within a measured section or borehole data, the resulting tally matrix will have matching row and column sums which are identical or nearly identical. Such a transition count matrix is said to have marginal homogeneity. If it is found that the tally matrix lacks marginal homogeneity, even when tested by a chi-square test, the results should be discarded and the procedure for collecting the data carefully reviewed. The most likely cause of failure is a bias through compilation of counts from a large number of disjoint borehole data. If the transition tally matrix has marginal homogeneity a simple chi-square test for symmetry model is sufficient to determine if the sedimentary successions follows a reversible or a nonreversible Markov sequence. If the sequence is nonreversible then the forward and backward succession are definitely different and such a distinction cannot be made if the sequence is reversible. A new method Kolmogorov criterion is introduced for checking transition matrices of reversible or nonreversible Markov process without requiring knowledge of the stationary probabilities of observed transition probability matrix.

The methods should be considered as a means of achieving unbiased results or, constraining geological interpretation of relationship between observable lithofacies that are more often very difficult to perceive physically from stratigraphic or borehole sections. Hopefully, this research will encourage and enrich the concept of coal and hydrocarbon exploration and development, especially in stratigraphy cases

**Acknowledgements**

We are grateful to statisticians and geoscientists who pioneered the application of stochastic processes and subsequent researchers for successful application in subtle problems. Prof. R. N. Hota has been kind enough to draw the beautiful illustrations. The author is thankful to Dr. Foroogh Abasaghi, Firduas University, Iran kind enough for providing published papers used in this study. Authors hereby declared the NO generative AI technologies and text-in-image generators have been used during preparation of manuscript. Last but not least we are thankful to the anonymous referees for their suggestions and comments on the previous version of the paper to improve it.

**COMPETING** **INTEREST**

The research does not receive any support from the funding agencies. The authors have declared that no known competing interest exist.

**References**

[1] Holbrook, J. M. and Miall, A.D. Time in the rock: a field guide to interpreting past events and processes from siliciclastic stratigraphy. Earth Science Reviews, 2020. **203,** 103121.doi.org/10.1016/.earscirev.2020.103121

[2]. A. D .Miall . Principles of Sedimentary Basin analysis. 3rd edition, Springer Sciences Business. 2013. 605p.

[3]. Xu, G. and Haq, B.U. Seismic facies analysis: Past, present and future. Earth Science Reviews. 2022. **224,** 103876. Doi.org/10.1016/earscirev.2021.103876

[4]. Chapman, W. A. L., and Finnegan, N. J. The signature of climate in fluvial suspended sediment records. Journal Geophysical Research: Earth Surface,2024. **129**. doi.org/10.1029/2023/F007429

[5]. Middleton, G. V. Johannes Walther’s Law of the Correlation of Facies. Bulletin Geological Society America. 1973. **84**, 979-988.

[6]. Sengupta, S. M. Introduction to Sedimentology. 3rd.Edition. Oxford and IBH Publishing Company, New Delhi. 2018. 314p.

[7]. Carr, T. R. Log-Linear models, Markov chain and cyclic sedimentation. Journal Sedimentary Petrology. 1982. **52**, 905-912.

[8]. Casshyap, S.M. and Tewari, R.C. Fluvial models of Lower Gondwana coal measures of Son-Mahanadi and Koel-Damodar valley basins. International Association of Sedimentologists, Special Publication, 1984, **7**, 121-143.

[9]. Tewari, R.C., Hota R. N. and Maejima W. Fluvial architecture of early Permian Barakar rocks of Korba Gondwana basin, eastern-central India. Journal Asian Earth Science, 2012, **52**, 43-52.

[10]. Khan, Z. A. Tewari, R. C. and Hota, R. N. Facies analysis, Markov model and linking of sub-environments in the early Permian Barakar coal measures of Godavari Gondwana basin of southeastern India. Journal Geological Society India. 2020, **95**, 599-608.

[11]. Davis, J. C. Statistics and Data Analysis in Geology. John Wiley & Sons. 2002. 638pp.

[12]. Hota, R. N. Adhikari, P. C. and Maejima, W. Cyclic sedimentation of the Barakar Formation, Singrauli coalfield, India: Statistical assessment from borehole logs. Open Journal Geology. 2012. **2**, 1-13.

[13]. Setiadi, D. J. Haryanto, I, Muljana, B. and Alam, S. A combination of embedded Markov chain and Binomial probability as a tool in sedimentary facies analysis. Indonesian Journal Geosciences. 2022. **9**, 291-302.

[14]. Selley, R. C. Studies of sequence in sediments using a simple Mathematical device. Quarterly Journal Geological Society London. 1979. **125**, 557-581.

[15]. Krumbein, W. C. FORTRAN IV computer program for simulation of transgression and regression with continuous time Markov models. Computer Contribution Geological Survey Kansas. 1967. **26**, 38p.

[16]. Krumbein, W. C. and Dacey, M. E. Markov chain and embedded chains in Geology. Mathematical Geology. 1969. **1**, 79-96.

[17]. Powers, E. D. and Easterling, G. R. Improved methodology for using embedded Markov chains to describe cyclical; sediments. Journal Sedimentary Petrology. 1982. **52**, 913-923.

[18]. Le Roux, J. P. Spreadsheet procedure for modified first order embedded Markov analysis of cyclicity in sediments. Computer and Geosciences. 1994. **20**, 17-22.

[19]. Harbaugh, J.W. and Bonham-Carter, G.F. Computer simulations in Geology. Wiley, New York, 1970. 575p.

[20]. Rolke, W. A. Continuous time Markov process as a Stochastic model for sedimentation. Mathematical Geology. 1991.**23**. 297-304.

[21]. Lie, W. D. A fixed-path Markov chain algorithm for conditional simulation of Discrete spatial variables. Mathematical Geology**,** 2007, **39,** 159-176.

[22]. Tewari, R. C. Stratigraphic simulation of late Paleozoic coal bearing cycles in Gondwana basins of eastern India using continuous and time Markov model. Journal Geosciences Research. 2017. **2,** 1-8.

[23]. Gingerich, P. D. Markov analysis of cyclic alluvial sediments. Journal Sedimentary Petrology. 1969. **39**, 330-332.

[24]. Billingsley, P. Statistical methods in chains. Annals Mathematics Statistics, 1961, **32**. 12- 40.

[25]. Hattori, I. Entropy in Markov chains and discrimination of cyclic patterns in lithologic successions. Mathematical Geology. 1976. **8**, 477-497.

[26]. Cant, D. J. and Walker, R. G. Development of a braided fluvial facies model for the Devonian Battery Point sandstone, Quebec. Canada Journal Earth Science. 1976. **13**, 102-119

[27]. Harper, C. W. Improved methods of facies sequence analysis. In: Facies Models (Ed) Walker, R, G. Geological Association Canada, Toronto. 1984. 11-13.

[28]. Casshyap, S. M. Kreuser, T. and Wopfner, H. Analysis of cyclical sedimentation in the lower Permian Mchuchuma coalfield (South-West Tanzania). Geologische Rundschau. 1987. **76**, 869-883.

[29]. Casshyap, S. M. and Khan, Z.A. Discrete Markov analysis of the Permian coal measurtes of East Bokaro basin, Bihar. Indian Journal Earth Sciences, **9**, 99-110

[30]. Sen, S. and Dey, J. Cyclic sedimentation in the Barakar Formation of the Karanpura field, marginal Gondwana basin, India. Journal Geological Society India, 2020. **95**, 293-300.

[31]. Johnson, K. R. and Cook, A. C. Cyclic characteristics of sediments in the Moon Island Beach Subgroup, Newcastle Coal measures, New South Wales. Mathematical Geology. 1973. **5**, 91-110.

[32]. Lever, H. Cyclic sedimentation in the shallow marine upper Permian Kennedy Group Carnarvon basin, Western Australia. Sedimentary Geology. 2004. **172**, 187-209.

[33]. Sarmah, R. K. Study of cyclic patterns and lithofacies variability by application of Markov chains and entropy analysis to Tikak Parbhat Formation, Makum coal basin, Northeast India. European International Journal Science and Technology. 2013. **2**, 41-49.

[34]. Strahaler, A. N. The Earth Sciences. Harper and Row, New York. 3rd. Edition. 2016. 681p.

[35]. Xu, H. and McCarthy, I. A. J. Markov chain analysis of vertical facies sequences using a computer software package (SAVFS). Courtmacsherry Formation, Southern Ireland. Computer and Geosciences.1998, **24**, 131-139.

[36]. Turk, G. Transition analysis of structural sequences-discussion. Bulletin Geological Society America. 1979, **90**, 989-991.

[37]. Walther, J. Einleitung in die Geologie als historische Wissenschaft. 1893. **1055**, Verlag Gustav Fisher.

[38]. Dalrymple, R. W. Interpreting sedimentary successions: Facies, Facies analysis and Facies models: In Facies Model 4th.Edition.Geological Association of Canada. 2010, 3-18. doi: 10.1002/9781444304091.ch.7.

[39]. Bridge, J. S. and Mackey, S. D. A revised alluvial stratigraphic model; In; Alluvial Sedimentology. (Eds) Marzo, M. and Pingdafabregas, C. Special Publication. International Association of Sedimentologists, 1993. **17**, 319-336.

[40]. Khan, Z. A. and Tewari, R. C. Analyzing sedimentary rocks to evaluate paleo-dimensions and flow dynamics of Permian Barakar River of Rajmahal Gondwana sub-basin of eastern India. In: Sedimentary Rocks and Aquifers:NewInsight.2022.129.doi:http://dx.doi.org/10.5772/intechopen.106994.

[41]. Singh, A. and Bhardwaj, B. D. Fluvial facies model for the Ganga River sediments, India. Sedimentary Geology. 1991. **72**, 135-146.

[42]. Sarma, J. N. Fluvial process and morphology of the Brahmaputra River in Assam. Geomorphology, 2005, **70**, 226-256.

[43]. Colombera, L. and Mountney, N. P. The lithofacies organization of fluvial channel deposits: A meta-analysis of modern rivers. Sedimentary Geology. 2019. **383**, 16-40.

[44]. Pettijohn, F. J. Sedimentary Rocks. CBS Publication, New Delhi. 2004. 628p.

[45]. Belt, E. S. Sakimoto, S.E.H. and Rockwell, B. W. A drainage diversion hypothesis for the origin of widespread coal beds in the Williston basin: Example from Paleocene strata, Eastern Montana. In: Coal Geology of Montana. (Ed.) Sholes, M. A. Montana Bureau Mines Geology, Special Publication.1992. **102**, 21-60.

[46]. Maejima, W. Tewari, R. C. and Hota, R. N. On the origin of Barakar coal-bearing cycles in the Gondwana basins of peninsular India. Journal Geosciences Osaka City University, Japan. 2008. **51**, 21-26.

[47]. Agresti, A. An introduction of categorical data analysis. John Wiley & Sons. 3rd. edition. New York. 2018.

[48]. Plackett, R. L. The Analysis of Categorical data. Charles Griffin, London. 1984. 206p.

[49]. Sharp, W. E. and Markham, T. Quasi-symmetry and reversible Markov sequence in sedimentary sections. Mathematical Geology. 2000. **32**, 561-579.

[50]. Khan, Z. A. and Tewari, R. C. Markov reversibility, quasi-symmetry, and marginal homogeneity in cyclothymiacs geological successions. International Journal Geo-informatics and Geological Sciences. 2021. doi:10.14445/23939206/IJGGS-V812P102,9-25.

[51.Tahata, K. Advances in quasi-symmetry for square contingency tables. Symmetry. 2022. **14**, doi.org/10.3390/sym 140510512-16.

[52].Caussinus, H. Contribution al’analyse statistique des tableaux de correlation. Annales de la Faculte des Science de l’Univetsite Toulouse. 1965. **29**, 77-183.

[53].Stuart, A. A test for homogeneity of the marginal distribution in a two way classification. Biometrika. 1955. **42**, 412-416.

[54]. Bhapkar, P. A note on the equivalence of two test criteria for hypothesis in categorical data. Journal American Statistical Association. 1966. **61**, 228-235.

[55]. Tewari, R. C. Singh, D. P. and Khan, Z. A. Application of Markov chain and Entropy analysis to lithologic succession- an example from the early Permian Barakar Formation, Bellampalli coalfield, Andhra Pradesh, India. Journal Earth System Science. 2009. **118,** 583-596.

[56]. Le Roux, J. P. Paleo-environmental interpretation of tabular sandstones in Beaufort Group of the Karoo basin, South Africa. South African Journal Geology. 1992. **95**, 171-180.

[57]. Khan, Z. A. Quasi-independence and entropy analysis of a late orogenic middle Siwalik molasses sequence in Kuluchaur area, Pauri Garhwal, India. Geoinformatics. 1996, **7**, 135-145.