Journal of Data Science **4**(2006), 1-20

On the Use of Geostatistical Cross-Association Method for Lithostratigraphical Correlation

Walid Abdolqader Saqqa and Mohammad Fraiwan Al-Saleh
Yarmouk University

Abstract: The aim of this paper is to determine the effectiveness of cross association in detecting the similarity between correlated geological columnar sections. For this purpose, *cross association* is used to compare several geological columnar sections which are arbitrarily selected from different localities in central and north Jordan. It turns out, for most of the study cases, that the sections which consist of the same rock units (formations) are statistically classified as similar (p-value $\ll .05$), while sections of different rock units (formations) are statistically classified as dissimilar (p-value $\gg .05$).

Key words: Chi-square test, cross association, geological sections, geostatistics, lithostratigraphical correlation, rock formation.

1. Introduction

Geologists get used to study rocks in outcrops or in drill cores. A sequence of sedimentary rocks may be divided up into a number of lithostratigraphical units usually of various sizes. Superimposed lithostratigraphical units are frequently distinguished from each other based on lithological properties and stratigraphical position. The formation is the fundamental unit of stratigraphy with definite lithological composition or a distinctive interbedded or intergraded succession of rock types that must be differ from the adjacent units above and below. The formations in a rock sequence named for some geographical localities that show typical exposures of these formations (known as type sections), are usually traceable or mappable from one exposure to another or from well to well in the subsurface (Krumbein and Sloss, 1963; Doyle *et al.*, 1994). Geologists can draw stratigraphical sections for several outcrops (or cores) in an area, and trace beds from section to section that finally leads to what is known as lithological correlation. This type of correlation basically demonstrates the equivalency of rock units across an area. The matter of correlation between different geological sections is relatively easier if the formation in the vertical rock sequence is not so greatly different in its lithology and thickness from one place to another. However, in many cases the lithological correlation meets with difficulties due to the fact of lateral change in bed thickness and lithology, missing of strata by erosion, lack of fossils, and tilting of strata. In such cases, the correlation between rock units becomes not so easy; it will be even difficult to designate the exact stratigraphical position of a rock column when it is compared with the type section of the same formation. In such cases, the use of geostatistics may be helpful. One possible approach that is used by geologists is the *cross association* where the applied data consist of two or more lithotype series of encountered in geological sections.

2. Aim of Study

The present study aims to attest the validity of the geostatistical cross association method in facilitating the exact lining-up of stratified rocks and the correlation between different geological sections. Great variations in lithology and thickness of rock units that may appear from one geologic exposure to another, and the obscurity in demarcating the exact boundaries between superimposed formations may call for applying the foregoing statistical method to resolve such problems.

3. Rock-Typing and Description of Data

The lithostratigraphical units used for statistical analysis, were arbitrarily selected from the exposed Upper Cretaceous (Cenomanian-Turonian) carbonate rock-type of the Ajlun Group (Quennell, 1951; Burdon, 1959) in central and north Jordan. The carbonate rock sequences of the Ajlun Group encompass five formations; these are from bottom to top: Naur (oldest), Fuhais, Hummar, Shuayb, and Wadi Es-Sir (youngest) (Masri, 1963). In fact, the location of boundaries between these formations becomes more difficult, and the thicknesses of rock units thin out in outcrops from north to south. Therefore, the Fuhais, Hummar and Shauyb formations in south Jordan are always studied as an integrated or undifferentiated unit under the name F/H/S formation. Ajlun Group primarily consists of limestone interbedded with dolomite, marl, and chert. Sedimentation of the Ajlun Group was during transgression of seawater in the Cenomanian-Turonian time, where Jordan substantially was a part of a broad epicontinental shelf of the Tethys and where most parts of it covered by seawater (Powell, 1988 and 1989; Abed, 2000).

Ajlun Group normally shows some lateral and vertical lithofacies changes as a result of change in sea-basin configuration. For example, the uppermost part of the Shuayb formation in west central Jordan witnesses lithofacies changes

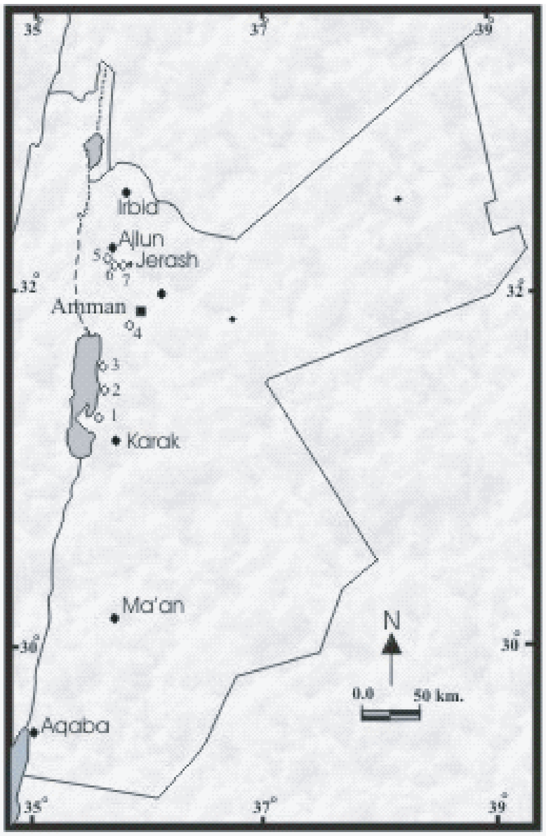

Fig. 1. Location map for investigated sections at areas from Ben Hammad (1), Mujib (2), Mukawir (3), El-Bahhath-Adassiya (4), Burma (5), Jarash (6), Inbeh (7).

whereby bedded-nodular gypsum and red-green claystone intercalate the predominating marl, marly limestone and limestone beds. The change in lithofacies, is a remarkable sign for variation in seawater level from deep to very shallow supratidal or coastal sabkha, unlike the case in north Jordan, where deep-water conditions prevailed (Abed and El-Hiayari, 1986). Further to the south, thin beds of gypsum can be found in the F/H/S formation (Powell, 1988 and 1989; Khalil, 1992). Calcrete soil-type coexists together with gypsum particularly in southwest Jordan (Ouran, 1996). Moreover, Abed and El-Hiayari (1986) observed some sandstone beds in the F/H/S formation in several locations in the south.

The thickness of Ajlun Group may exceed 600 m. in sections from north Jordan (Abdelhamid, 1995), decreases up to 450-350 m. around Amman and in

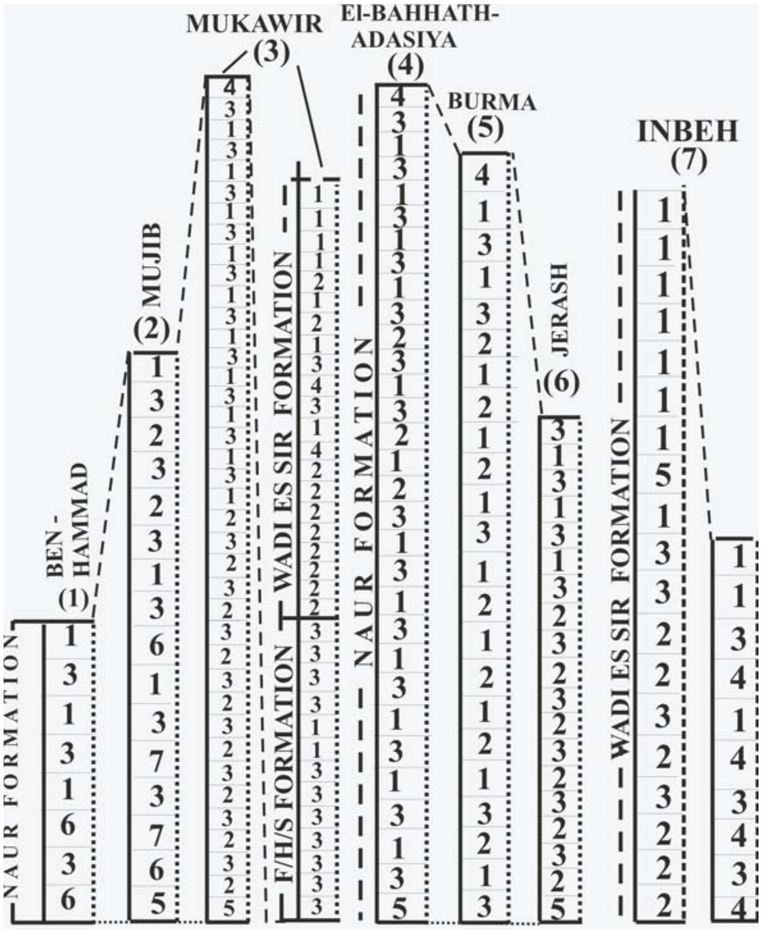

Fig.2. Graphic logs of the studied geological columnar sections with numerical labeling.

central parts of Jordan and diminishes to ±100 m. in the south and southeastern parts of the country (Abed, 2000).

4. Methodology

4.1 Selection of the geological columnar sections

To bring about *cross association* statistical study, we have selected a number of geological columnar sections from the Naur, F/H/S and Wadi Es Sir Formations in Ajlun Group. The columnar sections earlier described and reported by the Natural Resources Authority at different localities in north, central, and south Jordan(Abdelhamid, 1995; Makhlouf et al., 1996). We have firstly examined the validity of *cross association* by comparing together the sections built up of similar rock units. Secondly, we compared between columnar sections of different rock units. The studied sections (Figure 1) named as: Mujib (MUJ), Mukawir (MUK), Ben Hammad (BHD) (central Jordan), EL-Bahhath-Adassiya (BAH-AD), Jarash (JAR) , Inbeh (INB), Burma (BUR) (north Jordan). The graphic logs of the columnar sections are illustrated in nominal values instead of lithological description. (Figure 2). Types of lithology and their equivalent nominal values are illustrated in Table 1.

4.2. Principle of Cross Association

There are different techniques that can be used to analyze data consisting of a series of states. Since the data presented in this study are in nominal scale, statistical methods based on Pearson correlation coefficient cannot be used for the analysis, because these methods require data measured in a ratio scale. Thus, the available data can only be analyzed using the alternative *cross association* method, an approach that is commonly used by geologists. *Cross association* is an index that tends to measure the degree of similarity or correspondence (equivalence) between two segments or two sequences.

In this study, the two segments represent two geological columnar sections taken from two different localities. The two segments are of lengths n and m respectively, and that the variable of interest has k different nominal values, coded as $1, 2, \ldots, k$. To asses the degree of similarity between two sequences (sections), the nominal values in a given sequence are moved stepwise past the nominal values of a second sequence. At each step, the matching position, number of comparisons (the length of the overlapped segments) and the number of matches are recorded. Total number of matches at each matching position (N) is of a great interest. The *cross association index* (CAI) is the ratio of the number of matches to the length of the two overlapping segments. Assuming that the number of matches at position $i, i = 1, 2, \ldots, m + n - 1$, is y_i, then CAI is given by

$$CAI(i) = \frac{t}{\Delta},$$

where Δ is the length of the overlapping segments, which takes the values $1, 2, \ldots,$ $\min(n, m, i)$. Clearly, $CAI(i)$ is between zero and one. A large value of $CAI(i)$ is a similarity indication of the two segments. For example, if we have two sequences of length 4 each, such as 1,3,2,3 and 3,1,2,2; we shall expect the following situations:

Match position 1: 1,3,2,3
3,1,2,2

In this case $t = 0$ and $CAI(1) = 0$,

Match position 2 : 1,3,2,3
3,1,2,2

In this case $t = 0$ and $CAI(2) = 0$,

Match position 3: 1,3,2,3
3,1,2,2

In this case $t = 2$ and $CAI(3) = 2/3$,

Match position 4: 1,3,2,3
3,1,2,2

In this case $t = 1$ and $CAI(4) = 1/4$.

Match position 5: 1,3,2,3
3,1,2,2

In this case $t = 1$ and $CAI(5) = 1/3$.

Match position 6: 1,3,2,3
3,1,2,2

In this case $t = 0$ and $CAI(6) = 0$.

Match position 7: 1,3,2,3
3,1,2,2

In this case $t = 1$, and $CAI(7) = 1.0$.

Now, we are interested in testing the hypothesis:

$$H_0: \quad \text{The two sequences are not similar}$$

versus $H_1: \quad H_0$ is not true.

Meanwhile, one may ask: how large $CAI(i)$ should be to conclude that H_0 is not true. For answering the question, let us assume that we have two random segments (sequences); each has the same number of observations and are of same composition. Firstly, we must find out the expected total number of matches. At any position, we may either have a match with probability p or a mismatch with probability $1-p$. Thus, at any position, we have a Bernoulli trial with a probability of success $= p$. Let the values of the first sequence be $a_1, a_2, \ldots, a_k$, where a_i denotes the total number of times that state i occurs, $i = 1, 2, \ldots, k$; $\sum_{i=1}^{k} a_i = n$, and for the second sequence, the values are denoted by $b_1.b_2.\ldots, b_k$; $\sum_{i=1}^{k} b_i = m$.

Using some counting techniques, the total number of possible ways for filling, at random, any matching position with two values of k (one value is for sequence 1, the second value is for sequence 2) is mn. The total number of possible ways for filling any matching position with two identical values of k — where a match occurs, is $\sum_{i=1}^{k} a_i b_i$. Thus, the probability of a match between two sequences at any position is given by:

$$p^* = \frac{1}{mn} \sum_{i=1}^{k} a_i b_i$$

(For more details about counting techniques, see Scheaffer,1995; Ghahramani, 2000). At this point, we can say that the two investigated segments originate from two populations with unknown p. The H_0 and H_1 can be rewritten as

$$H_0: \quad p \leq p^*, \quad \text{versus} \quad H_1: \quad p > p^*.$$

In the foregoing example $a_1 = a_2 = 1, a_3 = 2; b_1 = 1, b_2 = 2, b_3 = 1; k = 3, m = n = 4$. Thus, the probability of a match at any position of comparison, for any two random sequences of the same compositions as those in the example, is $p^* = 5/6 = .3125$, while the probability of a mismatch is $1 - 5/6 = .6875$. If an overlapped sections are of length Δ, i.e., Δ is the total number of comparisons, then the expected number of matches is $E = \Delta p$. In the previous example, Δ can take the values $1, 2, 3, 4$. If $\Delta = 1$, then $E = .3125$. On the other extreme, if $\Delta = 4$, then $E = 1.25$. Note that the observed number of matches when $\Delta = 4$ for the two examined sequences is 1.

Given two segments, we want to determine whether the observed number of matches, at a given matching position is significantly large (compared to E so that the hypothesis H_0 is rejected, i.e., the two sections are comparable. The only need is to use a statistical test of significant.

Let O be the total number of matches for an overlapping segments of length Δ. Given the value of Δ, O is a sum of Δ Bernoulli trials with common probability of success p^*. However these trials are not independent because there is a match at position i that will reduce chances of getting a match at position j. In other word,

$$p^{**} = Pr(\text{a match at position } j \,|\, \text{ there is a match at position } i) \neq p^*.$$

Thus O is not a binomial random variable. However, the expected value of O continues to be Δp^*. It can be shown that the conditional probability p^{**} is given by

$$p^{**} = \frac{mn}{(m-1)(n-1)} p^* - \frac{k + \sum_{i=1}^{k} a_i b_i (a_i + b_i)}{mn(m-1)(n-1)p^*}.$$

As it can be seen from the above formula, p^{**} gets closer to p^* as m and n get larger. Thus, for fairly large values of m and n which are nearly valid in our case -and under some conditions, approximate distribution of O can be found in the literature under the null hypothesis. Serfling (1975) showed that the distribution of O can be approximated by the Poisson distribution. The approximation is very accurate when the trials are not strongly dependent. Thus the probability that O is larger than a given number can be approximated under the null hypothesis and hence the p-value of the test can be determined. The same argument of Serfling (1975) can lead to a binomial approximation and hence the χ^2-test can be used as an approximate test:

$$\chi^2 = \frac{(O-E)^2}{E} + \frac{(O'-E')^2}{E'},$$

again O is the number of matches for a given Δ, O' is the number of mismatches given by $\Delta - O$. Also, $E' = \Delta - E$. The approximate distribution of χ^2, under the null hypothesis is the well known χ^2-distribution with one degree of freedom. Yates correction may be applied to the χ^2-statistics, especially when the expected number of matches is small. This is common near the ends of chains. Yates correction calls for a subtraction of .5 from the absolute difference of the observed and expected number of matches. (See Davis, 1986, pp. 236-237.). The modified χ^2 statistic becomes:

$$\chi^2_Y = \frac{(O-E-0.5)^2}{E} + \frac{(O'-E'-0.5)^2}{E'}.$$

Large values of χ^2 or χ^2_Y indicate that there is some similarity between the two sequences (i.e., rejection of H_0), while small values of χ^2 or χ^2_Y indicate that the similarity between the two sequences is just as that for two independent random sequences of the same composition. H_0 is rejected if χ^2 or χ^2_Y is larger than a critical value taken from the χ^2-table (see for example, Brase and Brase, 2004) for a given level of significant α. The most popular value of α is 0.05; the corresponding critical value is 3.84.

In the next section, we will apply the above-mentioned statistical methodology in comparing the different pairs of geological columnar sections. Given a pair of sections we need to:

1. determine the position of best match between the two sections, and to obtain the value of Δ, as well as the observed number of matches, o.

2. approximate the conditional null distribution of O given Δ. In particular, we approximate the p-value as:

   $$p - \text{value } = P(O > o \,|\, \Delta, H_0) = 1 - \sum_{r=0}^{o} \frac{\lambda^r e^{-\lambda}}{r!},$$

   where $\lambda = E = \Delta p^*$.

Alternatively, we can obtain the χ^2_Y and compare it with the critical value.

5. Data Analysis

5.1. MUK vs. MUJ and BHD Naur formation (central Jordan)

In this study case, three columnar sections made of *Naur formation* are mutually compared: MUK, MUJ and BHD (central Jordan). The variable of interest (Lithotype) in the Ajlun Group sequence has seven values (states), labeled $1, 2, \ldots, 7$. In fact, not all the states are necessarily observed in the studied section; for example, state labeled 7 (*Dolomitic Sandy Marl*) is only observed in BHD section. MUK section considered as a type locality section; therefore, other sections are correlated with it. The numerical label or code for each state is given in Table 1 below:

Table 1: Main lithotypes of Ajlun Group and the corresponding numerical values.

State	Label
Limestone	1
Dolomite	2
Marl	3
Cherty Limestone	4
Sandy Limestone	5
Fine Sandstone	6
Dolomitic Sandy Marl	7

The summary of the observed counts for the three sections is given in Table 2.

Consider first MUK section vs. MUJ section. Using the notation of the previous section, we have $k = 7, n = 39, m = 16$. Hence,

$$P^*_{1,2} = \frac{1}{mn} \sum_{i=1}^{n} a_i b_i = \frac{157}{39 \times 16} = .2516.$$

Thus, under H_0, at any two overlapping segments of length Δ, the probability of a match is .2516 and the expected number of matches is $E = .2516\Delta$. Since $1 \leq \Delta \leq 16$, $.2516 \leq E \leq 4.0256$. Now, MUJ section has been moved by the MUK section one position at a time and compared at each match position. The best matching position is the 28-th position. The total number of matches at this position is 8, and the length of the overlapping segments is $\Delta = 16$, (See Table A1 in the appendix). Thus, the best match between the two sections is when the last state of the MUJ section (Sandy Limestone) is lined up with the 28-th state of MUK section (Dolomite). At this best matching position, the conditional expectation of the number of matches when $\Delta = 16$ is $E = .2516 \times 16 = 4.0256$, while the observed number of matches for the two sequences under consideration is $O = 8$. Thus, the observe value of χ^2_Y is 4.67. This value is larger than the critical value 3.84. Hence, the null hypothesis is rejected. At this overlapping position, the total number of matches is large enough to indicate that the two sections have an amount of similarity more than the similarity occurring when the two sections are any two independent randomly selected sections. To see how strong this conclusion is, one may calculate the p-value of the test, which is $Pr(\chi^2 > 4.67)$. Using the statistical package, MINITAB, the p-value for this test is found to be .0307, indicating that the evidence against H_0 is strong. Alternatively, we can approximate the p-value using the Poisson approximation. In this case, the

$$p\text{-value } = P(O > 8 \,|\, \Delta = 16, p = .2516) \approx 1 - \sum_{r=0}^{8} \frac{4.0256^r e^{-4.0256}}{r!} = .022.$$

Table 2: Summary statistics for MUK, MUJ and BHD Naur formation.

MUK section (type locality)		MUJ section		BHD section	
Lithotype	Count	Lithotype	Count	Lithotype	Count
1	10	1	3	1	3
2	9	2	2	2	3
3	18	3	6	3	0
4	1	4	0	4	0
5	1	5	1	5	0
6	0	6	2	6	2
7	0	7	2	7	0
Total	39	Total	16	Total	8

Consider next, MUK sectiob vs. BHD section. In this case, we have $k = 7, n = 39, m = 8$. Hence,

$$p^*_{1,3} = \frac{1}{mn} \sum_{i=1}^{k} X_i Y_i = \frac{57}{39 \times 8} = .1827.$$

The BHD section is moved by the MUK section one position at a time, and it is compared at each match position. The best position of 6 matches, occurred at the match position 24, where $\Delta = 8$, (See Table A1 in the appendix). As a result, the best match between the two sections, is when the last state of the BHD section (fine sandstone) is lined up with the 24-th state of MUK section (Dolomite). Given $\Delta = 8$, the conditional expectation of the number of matches is $E = .1827 \times 8 = .1.4615$, while $O = 6$. This gives an observed value of 15.04 for χ^2_Y, which is much larger than the critical value 3.84. Hence, the null hypothesis is rejected. At this overlapping position, the total number of matches is large enough to indicate that the two sections have an amount of similarity much more than the similarity that occurs when the two sections are any two independent randomly selected sections. The p-value for this test, is found to be .0001, indicating that the evidence against H_0 is very strong. Using the Poisson approximation, we have

$$p\text{-value } = P(O > 6 \,|\, \Delta = 8, p = .1827) \approx 1 - \sum_{r=0}^{6} \frac{1.4615^r e^{-1.4615}}{r!} = .0008.$$

If we compare between MUJ section vs BHD section , we have $k = 7, n = 16, m = 8$. Hence,

$$p^*_{2,3} = \frac{1}{mn} \sum_{i=1}^{k} X_i Y_i = \frac{19}{16 \times 8} = .1484.$$

When BHD section is moved by the MUJ section one position at a time, and it is compared with it at each match position, we can see that the best matching position is at position 12 with 4 matches and $\Delta = 8$. (See Table A1 in the appendix). This means that, the best match between the two sections is when the last state of the BHD section (fine sandstone) is lined up with the 12-th state of MUJ section (Dolomitic Sandy Marl). The conditional expectation of the number of matches - given that the length of the overlapped sections $\Delta = 8$, is $E = .1484 \times 8 = 1.1872$, and $O = 4$. The observed value of χ^2_Y is 6.1165 which is larger than the critical value 3.84. Hence, the null hypothesis is rejected. At this overlapping position, the total number of matches is large enough to indicate that the two sections have a degree of similarity above that similarity when the two sections are any two independent randomly selected sections. The p-value in this case is .0134, which is a strong evidence against H_0. The Poisson approximation of the p-value is .007.

As noted from the previous analysis, it appears that the Poisson approximation of the calculated p-values - when we use an *exact* test based on the test statistic o, is not in a conflict with the results of the chi-square test. Since, most users of statistics are familiar with the chi-square test, we recommend using it in problems similar to the one considered here in this paper.

5.2. AD vs. BUR and JAR Naur formation (north Jordan)

Here, we compared between three geological columnar sections in the Naur Formation: Adassiya (AD), Burma (BUR) and Jarash (JAR) sections. The summary of the observed counts for the three sections is given in Table 3.

Table 3: Summary statistics for AD, BUR and JAR Naur formation (north Jordan).

AD section		BUR section		JAR section	
Lithotype	Count	Lithotype	Count	Lithotype	Count
1	12	1	10	1	3
2	3	2	7	2	6
3	14	5	0	5	1
6	0	6	0	6	0
7	0	7	0	7	0
Total	31	Total	23	Total	19

Consider AD section vs. BUR section. In this case, we have $k = 7, n = 31, m = 23$. Hence, $p^*_{1,2} = .2973$. BUR section is moved by the AD section one position at a time. The best match occurs at the match position 24; the number of matches at this matching position is 10 and $\Delta = 23$. The conditional expectation of the number of matches given $\Delta = 23$ is $E = .2973 \times 23 = 6.8383$ and $O = 10$. The observed value of χ^2_Y is 1.8656. This value is smaller than the critical value 3.84. Therefore, the null hypothesis can't be rejected. At this overlapping position, the total number of matches is not large enough to indicate that the two sections have a degree of similarity over that similarity when the two sections are any two independent randomly selected sections. The statistical results assume that these two sections are relatively not similar; though, in nature this is not the case. The p-value of the test is, .172, indicating that the evidence against is relatively weak. The Poisson approximation of the p-value in this case is .09, which is also not significant at $\alpha = .05$.

But by comparing AD "Naur formation" vs. JAR "Naur formation", we got much better results. At first, $n = 31, m = 19$ and $p^*_{1,3} = .3073$. When JAR section is moved by the AD section one position at a time and compared with it at each match position, the best number of matches is 14 occurred at match position 22 where $\Delta = 19$. The conditional expectation of the number of matches given $\Delta = 19$ is $E = .3073 \times 19 = 5.8387$ while $O = 14$. Hence, the observed value of χ^2_Y is 15.753. This value is much larger than the critical value 3.84. Therefore, the null hypothesis is rejected. At this overlapping position, the total number of matches is large enough to indicate that the two sections have an amount of similarity much more than the similarity that occurs when the two sections are any two independent randomly selected sections of the same compositions. Thus, we may conclude that the two sections are very similar. The p-value of the test is .001, indicating that the evidence against H_0 is very strong. The Poisson approximation of the p-value is also .001.

5.3. MUK Naur formation (central Jordan) vs. AD Naur formation (north Jordan)

In this study case, we compared between the two sections: the type locality MUK Naur Formation, central Jordan and the AD Naur Formation, north Jordan. The summary of the observed counts for the two sections is given in Table 4.

Table 4: Summary statistics for the two sections: MUK and AD Naur formation.

MUK section		AD section	
Lithotype	Count	Lithotype	Count
1	10	1	12
2	9	2	3
3	18	3	14
4	1	4	1
5	1	5	1
6	0	6	0
7	0	7	0
Total	39	Total	31

In this case, we have; $n = 39, m = 31$. Hence, $p^*_{1,2} = .3317$. The best match, when AD section was moved by the MUK section one position at a time, occurred at match position 31. The total number of matches at this position is 17 with $\Delta = 31$. On the other hand, the conditional expectation of the number of matches given $\Delta = 31$ is $E = .3317 \times 23 = 10.2827$ while $O = 17$. Thus, the observed value of χ^2_Y is 6.2735. This value is larger than the critical value 3.84 and hence the null hypothesis should be rejected. At this overlapping position, the total number of matches is large enough to indicate that the two sections have an amount of similarity more than the similarity that occurs when the two sections are any two independent randomly selected sections. The p-value of the test is .0123, indicating that the evidence against H_0 is strong. The Poisson approximation of the p-value is .018.

5.4 MUK, F.H.S formation (central Jordan) vs. JAR, F.H.S formation (north Jordan).

In this section, we compared two columnar sections of the same rock type “ F/H/S formation”: the type locality MUK section (central Jordan) and JAR section (north Jordan). The summary of the observed counts for the two sections is given in Table 5.

Table 5: Summary Statistics for the two sections: MUK and JAR F/H/S formation.

MUK section		JAR section	
Lithotype	Count	Lithotype	Count
1	2	1	2
2	0	2	0
3	11	3	9
4	0	4	0
5	0	5	0
6	0	6	0
7	0	7	0
Total	13	Total	11

In this case, we have; $n = 31, m = 11$. Hence, $p^*_{1,2} = .7203$. when the JAR section is moved by the MUK section one position at a time and compared with it at each match position, the best match is of 9 matches occurred at match position 9 with $\Delta = 9$. The conditional expectation of $E = .7203 \times 9 = 6.4827$. The observed value of $\chi^2_Y = 4.2444$. The p-vaalue of the test is .0394, and the Poisson approximation of the p-value is .121. This is good indication that the evidence against H_0 is relatively weak.

5.5 MUK Wadi es-Sir formation (central Jordan) vs. INB Wadi es Sir formation (north Jordan).

In this case, we compared between two different sections made of Wadi-es-Sir limestone Formation: the type locality MUK section (central Jordan) and INB section (north Jordan). The summary of the observed counts for the two sections

Table 6: Summary statistics for the two sections: MUK and INB Wadi es-Sir formation.

INB section		MUK section	
Lithotype	Count	Lithotype	Count
1	11	1	7
2	6	2	10
3	7	3	2
4	4	4	2
5	2	5	0
6	0	6	0
7	0	7	0
Total	29	Total	21

is given in Table 6.

In this case, we have; $n = 29, m = 21$. Hence, $p^*_{1,2} = .2611$. The best match of 10 matches occurred at the match position 29, where $\Delta = 21$. The conditional expectation of the number of matches given $\Delta = 21$ is $E = 5.4831, O = 10$, and the observed value of χ^2_Y is 4.5648. This value is larger than the critical value 3.84 and hence the null hypothesis is rejected. At this overlapping position, the total number of matches is large enough to indicate that the two sections have an amount of similarity more than the similarity that occurs when the two sections are any two independent randomly selected sections. The p-value of the test is .0326, while the Poisson approximation of the p-value is .0248, indicating that the evidence against H_0 is moderate.

5.6 MUK Naur formation (central Jordan) vs. JAR F/H/S formation (north Jordan).

In this study case, We compared between two sections of different rock units: the type locality MUK Naur Formation and JAR F/H/S Formation. The summary of the observed counts for the two sections is given in Table 7.

In this case, we have; $n = 39, m = 12$. Hence, $p^*_{1,2} = .4274$. When JAR section is moved by the MUK section one position at a time and compared at each match position, the maximum number of matches is 7 occurred at match position 25 where $\Delta = 12$. On the other hand, the conditional expectation of the number of matches given that $\Delta = 12$ is $E = 5.1282$ and $O = 7$. The observed value of χ^2_Y is 1.1856. This value is much smaller than the critical value 3.84 and hence the null hypothesis is not rejected. Thus, at this overlapping position, the total number of matches is not large enough to indicate that the two sections have a degree of similarity above the similarity that occurs when the two sections are any two independent randomly selected sections. The p-value of the test is .2762, while the Poisson approximation of the p-value is .1471. Both values indicate that the evidence against H_0 is very weak.

Table 7: Summary statistics for Mukawir Naur formation and JAR F/H/S formation.

MUK section (Naur F.)		JAR section (F/H/S F.)	
Lithotype	Count	Lithotype	Count
1	10	1	2
2	9	2	0
3	18	3	10
4	1	4	0
5	1	5	0
6	0	6	0
7	0	7	0
Total	39	Total	12

5. Discussion and concluding Remarks

In this study, we statistically compared nine pairs of geological columnar sections using the cross association method. The comparison is based on the correlation between similar rock units (formations) in one hand and different rock units (formations) in the other hand. Statistical testing, based on cross association, was able to detect the similarity between pairs of the same rock unit “formation” in seven study cases out of eight study cases. The study case which does not show enough number of matches or a high degree of similarity is that between the AD ”Naur Formation” versus BUR “Naur Formation” in north Jordan. In the meantime, the test is able to detect a mismatch between MUK ”Naur Formation” (central Jordan) and JAR “F/H/S Formation” (north Jordan). Because of the high rate of success (89%) achieved by this study, geologists can use cross association to evaluate the rate of accuracy in the correlation between pairs of the geological sections, especially when the correlation becomes more difficult for the reasons formerly discussed in the introduction.

The statistical test used here is the total number of matches at the best matching position, O. Due to the dependency between trials, the exact null conditional

Table A1: Number of matches at different matching positions for MUK, MUJ and BHD Naur formation sections.

Matchpos.#	MUK vs. MUJ	MUK vs. BHD	MUJ vs. BHD
1	0	0	0
2	0	0	0
3	0	1	1
4	0	0	1
5	1	1	1
6	0	2	2
7	2	1	1
8	0	4	3
9	3	1	2
10	1	5	3
11	3	1	1
12	3	5	4
13	3	1	1
14	4	5	3
15	3	1	3
16	5	5	0
17	3	1	
18	6	5	
19	3	1	
20	6	5	
21	3	1	
22	6	5	
23	3	0	
24	6	6	
25	2	0	
26	7	4	
27	1	1	
28	8	2	
29	0	2	

continue to next page

Table A1 (continued): Number of matches at different matching positions for MUK, MUJ and BHD Naur formation sections.

Matchpos.#	MUK vs. MUJ	MUK vs. BHD	MUJ vs. BHD
30	7	1	
31	1	2	
32	5	1	
33	3	2	
34	4	1	
35	5	2	
36	3	1	
37	6	2	
38	2	1	
39	7	2	

distribution of O given the total number of comparisons, Δ, is not easy to tabulate. The p-value of the test is approximated using the chi-square distribution and the Poisson distribution; the two methods of approximations give similar results. Thus, we recommend using the more familiar chi-square approximation.

Acknowledgment

The authors wish to thank the editor and a referee for the useful comments and fruitful discussions that improve the quality of the paper.

Appendix

See Table A1.

References

Abdelhamid, Gh.. (1995). *The Geology of Jarash Area, Map Sheet (3154-I).* Natural Resources Authority, Geology Directorate, Geological Mapping Division, Bull. 30, Amman.

Abed, A. M. (2000). *The Geology, Environment and Water of Jordan.* The Publications of the Jordanian Geologists Association, Scientific Book Series, Amman, 571pp. (in Arabic).

Abed, A. M., El-Hiayari, M. (1986). Depositional environment and paleogeography of the Cretaceous gypsum horizon in west-central Jordan. *Sedimentary Geology* **47**, 109-123.

Brase , C. H. and Brase, C. P. (2004). *Understanding Basic Statistics.* Haughton Mifflin Company.

Burdon, D. J. (1959). *Handbook of the Geology of Jordan* (to accompany and explain the three sheets of 1:250,000 Geological Map, Kingdom of Jordan). Government of the Hashemite Kingdome of Jordan, Benham, Golchester.

Davis, J. C. (1986). *Statistics and Data Analysis in Geology.* John Wiley.

Doyle, P., Bennett, M. R., Baxter, A. N. (1994). *The Key to Earth History, An Introduction to Stratigraphy.* John Wiley and Sons.

Ghahramani, S. (2000). *Fundamentals of Probability.* Prentice Hall.

Khalil, B. M. (1992). The Geology of the Ar-Rabba Area (Map Sheet No. 3152 IV). *Bulletin 22, NRA, Mapping Project.*

Krumbein, W. C., Sloss, L. L. (1963). *Stratigraphy and Sedimentation.* W. H. Freeman and Company.

Makhlouf, I., Abu-Azzam, H., Al-Hiayrai, A. (1996). *Surface and Subsurface Lithostratigraphic Relationships of the Cretaceous Ajlun Group in Jordan.* Natural Resources Authority, Geology Directorate, Subsurface Geology Division, Subsurface Geol. Bull. 8, Amman.

Masri, M. (1963). The geology of the Amman-Zerqa area. Central Water Authority, Amman, Jordan (Unpublished Report).

Ouran, N. (1996). Petro-microfacies analysis and sequence stratigraphy of the Upper Cretaceous Ajlun Group, SW Jordan. Msc. Thesis, University of Jordan.

Powell, J. H. (1988). *The Geology of Karak; Map Sheet No. 3152 III*, Bulletin 8, NRA, Geology directorate, Amman, 171pp.

Powell, J. H. (1989). *Stratigraphy and Sedimentation of the Phanerozoic Rocks in Central and South Jordan. Part B Kurnub, Ajlun and Belqa Groups*, Bulletin 11, NRA, Geology directorate, Amman, 130pp.

Quennell, A. M. (1951). The geology and mineral resources of (former) Transjordan. *Colonial Geology and Mineral Resources, London* **2**, 85-115.

Scheaffer, R. L. (1995). *Introduction to Probability and its Applications.* Duxbury Press.

Serfling, R. J. (1975). A general Poisson approximation. *Annals of probability* **3**, 726-731.

Received Auggust 4, 2004; accepted November 25, 2004.

Walid Abdolqader Saqqa
Department of Earth and Environmental Sciences
Yarmouk University, Irbid, Jordan
wsaqqa@yu.edu.jo

Mohammad Fraiwan Al-Saleh
Department of Statistics
Yarmouk University, Irbid, Jordan
m-saleh@yu.edu.jo