RASC FORTRAN 77 COMPUTER PROGRAM
FOR RANKING AND SCALING
OF BIOSTRATIGRAPHIC EVENTS

M. Heller
W.S. Gradstein
F.M. Gradstein
F.P. Agterberg
S.N. Lew

November 1985

$25.00

R A S C

The Agterberg - Gradstein method for Ranking and Scaling

- an explanation of the methods and tests

- a manual for the implementation of the
  RASC computer program

- program input for a sample run

M. Heller   —   M. Heller Consultants, 2316 Maynard
                        Street, Halifax, N.S.

W.S. Gradstein   —   ECO - Science Interpretation Service,
                        26 Locke Street, Bedford, N.S.

F.M. Gradstein   —   Atlantic Geoscience Centre, Bedford
                        Institute of Oceanography, Dartmouth,
                        N.S.

F.P. Agterberg   —   Geological Survey of Canada, 601 Booth
                        Street, Ottawa, Ont.

S.N. Lew   —   Geological Survey of Canada, 601 Booth
                        Street, Ottawa, Ont.

TABLE OF CONTENTS

P U R P O S E

The purpose of this guide is to explain the Agterberg-Gradstein method for Ranking and Scaling of biostratigraphic events and to provide detailed instructions for its use. The actual program listing is stored on floppy discettes at the back of this manual.

The RASC (RAnking and SCaling) method and the RASC computer program, written in Fortran IV language, are to be used for:

(1) The determination of the average ranking of elements (i.e. stratigraphic events) in different samples (i.e. geological sections) from one assumedly homogeneous population (i.e. geographic region) along an irreversible scale (i.e. a relative timescale). This part is called Ranking.

(2) The determination of the average relative distance between these elements along that scale. This part is called Scaling.

(3) The detection of heterogeneity in the data by the testing of each sample (geological section) for "normality".

A biostratigrapher can use the program to construct range-charts of average ranges and zonations based on assemblage zones.

G E N E R A L   I N T R O D U C T I O N

The RASC method and program have been developed for use in biostratigraphy, the science that uses fossils to determine the geological age of rocks. Biostratigraphy is based on Darwinian evolution principles: In time biological taxa appear, reach a peak of abundance, and disappear, to make place for other taxa.

Ideally, the biostratigrapher likes to follow the Darwinian evolution (lineage) of one taxon. However, the fossil record, as found in sampled outcrops or in wells, never produces a long, continuous set of data. Strata might be geologically reworked or may sometimes go undetected. Their fossils might be absent (palaeo-ecological factors, high sedimentation rate) or be poorly preserved.

To alleviate these problems, the modern biostratigrapher tends to work in a quantitative way, generally using those stratigraphic events that are most common in the region studied. However, the varying thickness of stratigraphic units (strata representing a certain period of time) precludes the use of fossil information from different geological sections in one region for this quantitative approach. Correlation between sections calls for a relative scale. One way to obtain a relative scale is by averaging the sequences of the events in the individual sections.

RASC has been developed to handle the abundant information needed for this quantitative approach.

The program is based on a method initially proposed by Hay in 1972. It considers how often each common event (each event that occurs in at least a specific number of sections) occurs above, simultaneous with, or below each other common event in all geological sections (wells, outcrops) in one assumedly homogeneous region. By regrouping these relative orders (ranks) RASC eventually produces a true average sequence for the events. Since this average sequence can also be thought of as an "optimum sequence" the latter term is often used for the result of the 'ranking' procedure.

An event ranked in this way might be the first occurrence in time of a certain fossil. Another event might be the last occurrence in time of that same fossil. The period between the first occurrence and the last occurrence of a fossil is called the fossil's stratigraphic range. With the above ranking exercise a sequence of average ranges can be determined. RASC can thus also be used to produce an average range chart.

In practice the biostratigraphic characterization of a certain region does not only consist of an average sequence of single events or of an average range chart, but also of a sequence (zonation) of "assemblage zones", an assemblage zone being a body of strata characterized by an assemblage of fossils.

It is possible to use the average sequence of stratigraphic events to determine a sequence of fossil-assemblages along the time-scale. To detect these fossil-groupings it is necessary to estimate distances between the events, which can be done with the use of "cross-over frequencies".

Events tend to change places with one another in different sections. This is especialy true for benthic fossils, whose occurrences depend greatly on the local palaeo-environment, in addition to their stage in evolution. RASC uses this crossing-over to determine distances. It should be kept in mind that RASC distance estimates can thus not be made when crossing-over does not occur.

It is assumed that the more often the events change places, the closer together they are on the time-scale. This assumption, that cross-over frequencies and relative distances are related, is often used in other science disciplines.

It has turned out to be possible to detect clusters of events along the time-scale in this way, and extensive applications have shown that these clusters are equivalent to the above mentioned fossil-assemblages and that RASC can indeed be used to construct a zonation based on assemblage zones, (Gradstein and Agterberg 1982; Williamson, in press).

To relate the results of RASC to more traditional methods and to the existing literature, the program also provides for a "Unique Event Option", which ranks indicator

fossils and marker horizons in the average sequence, even if
these unique events were considered to be too rare to parti-
cipate in the ranking.

The third part of RASC is the normality test, which
consists of a series of statistical tests for the event
positions in each individual section in relation to the
optimum sequence. These tests have been found necessary
because biostratigraphic methods are subject to sample con-
tamination and error through fossil misidentification.

T H E   R A S C   M E T H O D S   A N D   T E S T S

The RASC program carries out the algorithms for the following methods and tests: Ranking, Scaling (with Marker Horizon and Unique Event Options), and Normality Testing. The methods and tests will be explained in this part of the guide. The original statistical and mathematical explanation can be found in Agterberg and Nel (1982a and b). A thorough evaluation of the method and results are in Gradstein et al. 1985.

A.   RANKING

(1)   The method

The ranking procedure employed in RASC, is based on a revised version of a method originally developed by Hay (1972). A so-called presorting option has been developed, which shortens the procedure considerably, and provisions have been made to alleviate the problem of "cycling events". As opposed to Hay's method, RASC considers simultaneous events as well.

The RASC procedure starts with the construction of a cumulative order matrix based on sequences of events observed in the individual sections. Only those events are considered that occur in at least k_c wells (common events). k_c is a threshold parameter and has to be set prior to ranking (see the appropriate item in "the implementation of the RASC program").

RASC lists all participating events in the margins of the cumulative order matrix, in the rather arbitrary sequence in which they have actually been identified. It is the objective to reorder these margin-events until their sequence averages all sequences found.

The elements of the matrix are scores, indicating in how many sections one event precedes another event. When, for example, event A precedes event B in 1 out of 6 sections, and B precedes A in 4 out of 6 sections, the matrix looks as follows:

	A	B
A	X	1
B	4	X

When A and B occur simultaneously on the 6th section, each score receives an extra 0.5:

	A	B
A	X	1.5
B	4.5	X

Thus, when a score S_{AB} (or more general s_{ij}) in the upper right triangle is lower than a score s_{BA} (or more general s_{ji}) in the lower left triangle, this means that B(j) precedes A(i) more often than A(i) precedes B(j). In the example the events in the margins are listed in the order in which they were found in the one section in which A precedes B. From the matrix-scores it now appears that the sequence A–B is not the average sequence. Only when s_{ij} in the upper triangle is higher than s_{ji} in the lower triangle is the sequence of the events in the margins the true average one. In practice, with many events, many scores s_{ij} in the upper triangle are lower than their counterparts s_{ji} in the lower triangle, and a regrouping of events is required, until all scores in the upper triangle are higher than, or equal to their counterparts in the lower triangle. Only then are the events in the margins ranked according to their average sequence.

For the first step, <u>the presorting option</u>, RASC derives a simplified matrix A from the cumulative order matrix S. The elements (a_{ij}) of this matrix are based on the scores s_{ij} in the S-matrix in the following way:

When $s_{ij} > s_{ji}$, $a_{ij} = 1$

When $s_{ij} < s_{ji}$, $a_{ij} = 0$

When $s_{ij} = s_{ji} \neq 0$, $a_{ij} = 0.5$

A simple example of a cumulative order matrix (S-matrix) and a resulting A-matrix is given below:

		A	B	C
	A	X	1.5	5
S-matrix:	B	4.5	X	5.5
	C	0	1.5	X

		A	B	C	Sums	Presort Rank Number
	A	X	0	1	1	2x1/2-0=1
A-matrix:	B	1	X	1	2	2x2/2-0=2
	C	0	0	X	0	2x0/2-0=0

When all elements of each row of the A-matrix are added, and the sums are ranked in descending order, the events approach the average sequence.

In practice RASC does not just determine the sums of the rows, because, when events i and j never occur together in a section, $s_{ij} = s_{ji} = 0$ and $a_{ij} = 0$, which is a different situation from $s_{ij} < s_{ji}$, $a_{ij} = 0$.

In order to distinguish between these two situations RASC determines how often the zeros in the ith row of the S-matrix correspond to $s_{ij} = s_{ji} = 0$. When there are b_i of these zeros in a row of (N-1) events the Presort Rank Number of event i is determined as follows:

$$\frac{(N-1) \times (\text{sum of all } a_i\text{-scores})}{(N-1)-b_i} \quad \text{(see A-matrix above)}$$

The next step, the matrix-permutation, is based on Hay's method. RASC lists for this step again the events by their sections, but now in a new sequence based on their Presort Rank Numbers, and constructs a new cumulative order matrix with the pre-sort sequence of events in its margins

- 7 -

(like B-A-C in the example above). Only those events are
used in this matrix that occur together in at least m_c
wells. If they occur together in fewer wells their relative
positions are not considered relevant. For example, if the
number of times, say r_{ij}, that two events i and j occur
together in a section is such that $r_{ij} < m_c$ then all
r_{ij} scores are set equal to 0 and are ignored in the
matrix-permutation. m_c has to be set prior to ranking
(see implementation of the RASC program) and should be
smaller than k_c.

All scores (s_{ij}) of this new matrix S are compared to
their counterparts (s_{ji}) and whenever a score s_{ij} in the
upper triangle is lower than its counterpart s_{ji} in the
lower triangle, the row i is interchanged with the row j and
column i is interchanged with column j.

This procedure is repeated until $s_{ij} > s_{ji}$ for all
elements in the upper triangle. The events in the margin
are then ranked according to their average sequence. In
future explanations this average sequence is referred to as
the "optimum sequence".

(2) Cycles in the matrix-permutation

In some cases the matrix-permutation will not produce a
solution. Cycles will occur. When, for instance, events A,
B and C get involved in a cycle, this means that A precedes
B more often than B precedes A and that B precedes C more
often than C precedes B, and that C precedes A more often
than A precedes C (A-B-C-A).

Cycles like these are fairly common and can be inter-
preted as follows:

(1)   A, B, and C occur simultaneously.
(2)   The data used is so patchy (especially when the
      k_c and m_c - values are very low), that no
      average sequence can be established.

Since the matrix-permutation cannot be carried through
when cycles occur, the cycles present a problem. In preced-
ing efforts (cf. Hay 1972) the cycling events were ignored
all together. However, the group of cycling events is also
stratigraphic information and should be included in the
result.

RASC starts the matrix-permutation, notes when a cycle occurs and which events are involved (cycles of three or four or even five or six events have been found in past uses of RASC), and consequently determines which two of the cycling events have the smallest difference in relative order $(s_{ij} - s_{ji})$. RASC zeros the scores of these two events, which breaks the cycle. If there is more than one pair of events that has the smallest difference in relative order ($s_{ij} - s_{ji}$ can be the same for several pairs of events), only one of those pairs is zeroed.

Once the cycle has been broken, the scores of the participating events can be calculated as before and these events will be included in the optimum sequence.

(3) Optimum sequence

The final solution for ranking appears in the RASC output directly after the listings (numerical and alphabetical) of all events of the input. It is presented under the heading "Optimum Sequence". Each event in this table receives a "range", which will be explained below. A copy of the Optimum Sequence of the sample run of Chapter V is given in Table 1.

(4) Ranges of the events in the optimum sequence

Each event is given a range in the average (optimum) sequence. This range should not be confused with the stratigraphic range of a taxon -- it only indicates that an event i has been found between the events of the range (h-k). Example: Event 12 of the sample run has position 9 and occurs between positions 8 and 10, but event 11 with position 10 occurs somewhere between positions 9 and 12. The second example means that the order relation of event 11 (position 10) with event 14 (position 11) has not been established, but that 11 does follow 12 (position 9) and precede event 15 (position 12) (see Table 1).

The detection of the range of an event in relation to other events takes place in the last S-matrix of the matrix-permutations, which is called the "final order relation matrix". If there are scores s_{ij} in the upper triangle that are:

(1)   ignored, because $r_{ij} < m_{c1}$;
(2)   zeroed, because they were cycle participants;
(3)   "tied", which means that $s_{ij} = s_{ji}$ and events
       i and j are simultaneous in the average sequence;

one has not determined whether j has a position either above
or below i and vice versa. The position of j is then part
of the range of i and the position of i is then part of the
range of j.

It should be stressed that the event-positions and
their ranges do not have a statistical meaning. The
positions are average, relative positions and their ranges
only indicate if an average sequence can be found between
one event and two or more of its neighbours.

B. SCALING

The scaling part of RASC gives statistical values to the distances between the events in the optimum sequence. These distance estimates result in a slightly different sequence (scaled optimum sequence). The scaled optimum sequence is graphically presented in a series of dendrograms that record the "distances" between events along a relative "time" scale.

Several dendrograms appear in the RASC output. Only the dendrogram that appears after the normality test tables contains the unique events, if they were included in the input. If marker horizons were included in the input they appear in all dendrograms (see Marker Horizons and Unique Events).

The RASC scaling method uses so-called cross-over frequencies (how often events change position with one another in different geological sections.), and is based on four assumptions (see below). It includes direct estimates, unweighted indirect estimates, and weighted indirect estimates of the distances between events, the construction of dendrograms, and a final reordering option.

(1) Cross-over frequency

The "cross-over frequency", F_{AB}, represents the ratio of sections in which A occurs above B. When A occurs above B in 9 out of 10 sections $F_{AB} = 0.9$ and the "distance" between A and B is considered to be large. When A occurs above B in 6 out of 10 sections and below B in 4 out of 10 sections, the "distance" is considered to be small. When $F_{AB} = 0.5$ the "distance" is 0, since A occurs as often above B as below B. When $F_{AB} < 0.5$, A and B are reversed and the "distance" is negative.

RASC uses the final order relation matrix of the ranking part of RASC to determine the cross-over frequencies. This implies that scaling has to be preceded by ranking.

(2) Theoretical assumptions

For the statistical method used in scaling as carried out by RASC, the following assumptions are made.

Assumption 1

The frequencies of the positions x_A and x_B etc. each, satisfy normal random (Gaussian) distributions with a mean of EX_A and EX_B etc. respectively, and a variance of σ^2. The Graphic representation of this assumption can be found in Figure 1.

Assumption 2

The frequency of the distance between two positions (x_B-x_A or d_{AB}) also satisfies a normal distribution, written as $f(d_{AB})$, which has a mean of EX_B-EX_A or Δ_{AB} and a variance of $2\sigma^2$. It is the Δ_{AB}, the mean distance between two events, or the distance between two events in the average sequence, that RASC will calculate a value for.

Assumption 2 follows directly from Assumption 1 (Fig. 1).

Assumption 3

The frequency of A occurring above B (F_{AB}) equals the statistical probability that A occurs above B ($F_{AB} = P_{AB}$). P_{AB} is then also the probability that $d_{AB} \geq 0$ ($P_{AB} = P(d_{AB} \geq 0)$). However, if P_{AB} were the probability that $d_{AB} \leq 0$ (Fig. 2), then the parameters Δ_{AB} and σ^2 of the normal distribution of d_{AB} could be directly calculated from the z-value:

$$z = \frac{d_{AB} - \Delta_{AB}}{\sigma\sqrt{2}} \rightarrow z = \frac{0 - \Delta_{AB}}{\sigma\sqrt{2}} = -\frac{\Delta_{AB}}{\sigma\sqrt{2}}$$

However, since $P_{AB} = P(d_{AB} \geq 0)$, and since it is known that:

$$P(d_{AB} \leq 0) = 1 - P(d_{AB} \geq 0)$$

the z-value (z_{BA}) for the probability that $d_{AB} \leq 0$ is the negative of the z-value (z_{AB}) for the probability that $d_{AB} \geq 0$ (P_{AB}) $\rightarrow -z_{BA}$. It follows that for P_{AB}

$$z_{AB} = \frac{\Delta_{AB}}{\sigma\sqrt{2}} \rightarrow \Delta_{AB} = \sigma\sqrt{2} \cdot z_{AB}$$
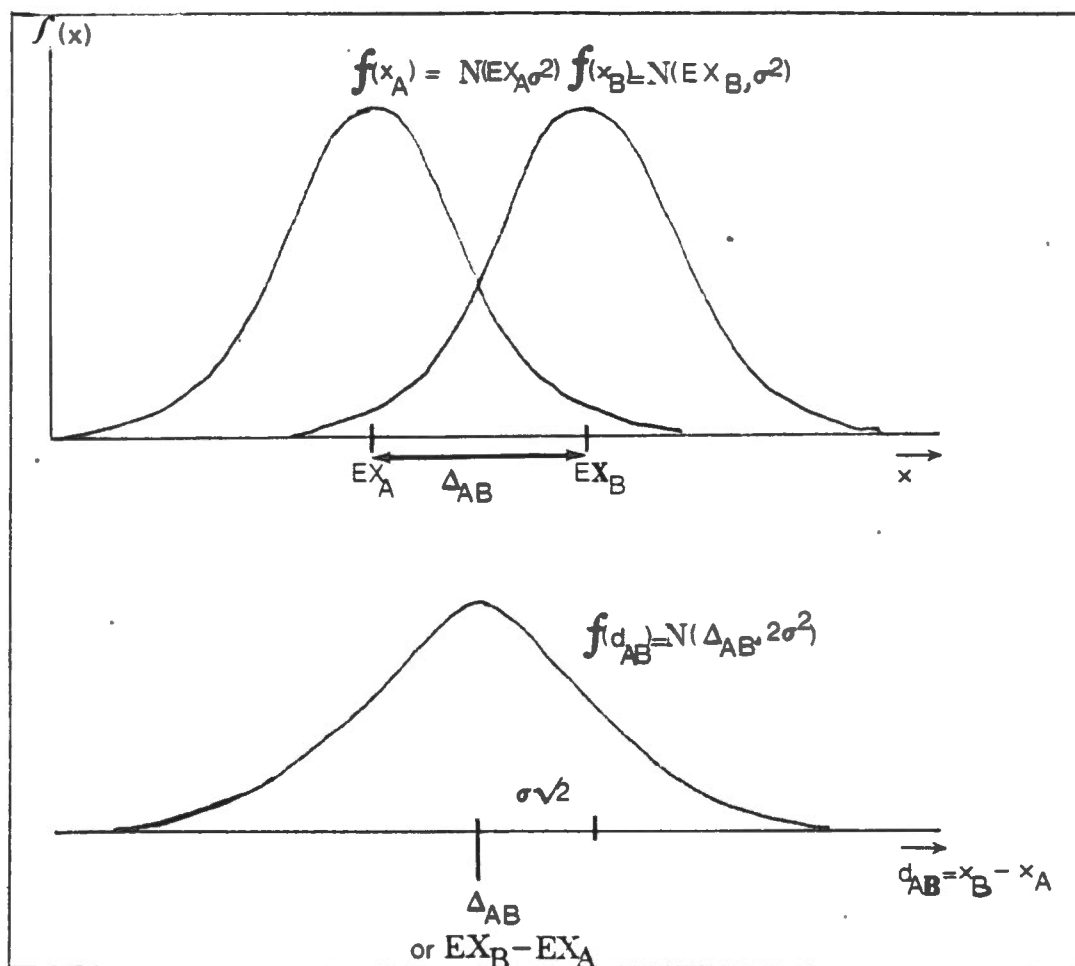
Figure 1. Graphic representations of assumptions 1 and 2.
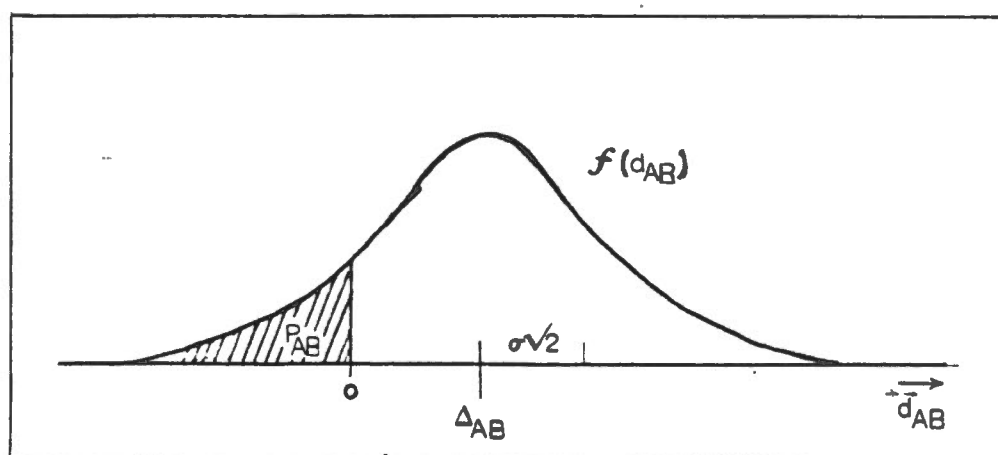
Figure 2. Graphic representation of P_{AB} being the probability that $d_{AB} \leq 0$, which is the opposite of assumption 3 ($P_{AB} = P[d_{AB} \geq 0]$).

Figure 3 gives the relationship between P_{AB}, z_{AB}, and $\frac{\Delta_{AB}}{\sigma\sqrt{2}}$

z_{AB} can be looked up in the tables for the cumulative standard normal distribution, P_{AB} being the cumulative frequency in %.
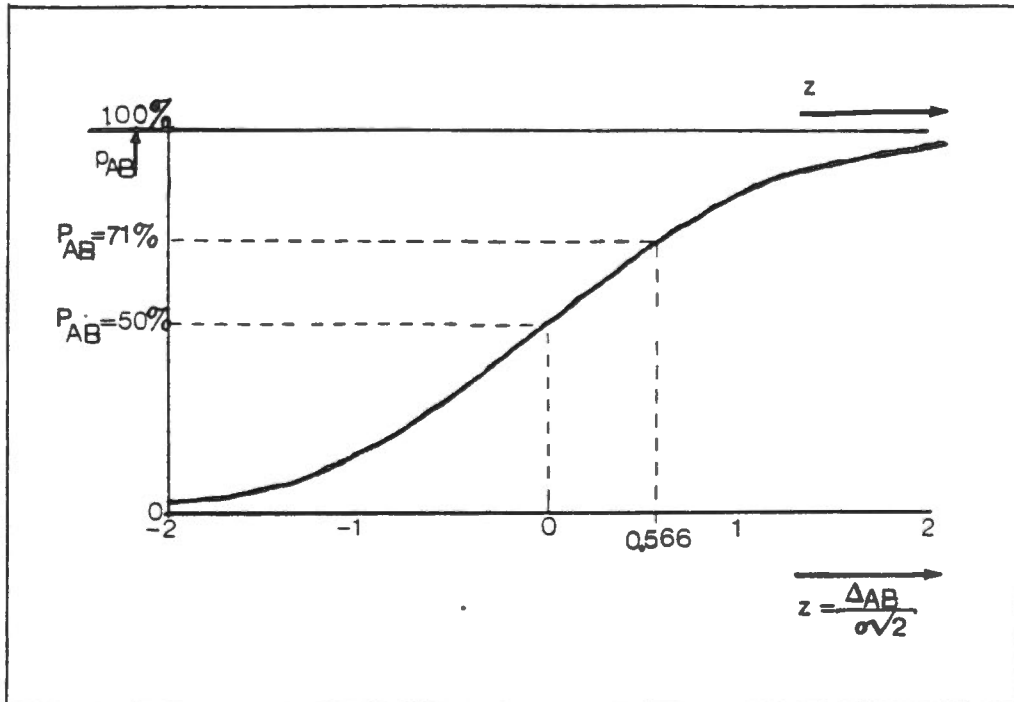
Figure 3. Cumulative frequency curve of a normal distribution. The curve indicates how all possible values for P_{AB} (P_{AB}) relate to the z-values of a normal distribution. The dashed lines indicate how one particular value of P_{AB} relates to one particular z-value, which equals, according to assumption 3, $\Delta_{AB}/\sigma\sqrt{2}$.

Examples:

When A occurs above B in 4 out of 8 sections
F_{AB} = 4/8 = 0.5 →→
P_{AB} = 50% see Fig. 3 and Table z = 0 → Δ_{AB} = 0
When A occurs above B in 5 out of 7 sections
F_{AB} = 5/7 → P_{AB} = 71% see Fig. 3 and Table z = 0.566 →→ Δ_{AB} = 0.566·$\sigma\sqrt{2}$

Assumption 4

$$\sigma^2 = 0.5$$

This assumption is wholly arbitrary, but can be made since the distance-scale is relative. It should be kept in mind that the distance-values eventually obtained are based on this arbitrary value, and that only differences between distances bear significance.

Since $\sigma^2 = 0.5 \rightarrow \sigma = \sqrt{0.5} \rightarrow \sigma \sqrt{2} = 1 \rightarrow \Delta_{AB} = z_{AB}$

(3) Direct distance estimates (Δ_{AB})

The scaling procedure as carried out be RASC, starts with the construction of a new matrix (F), which has the events in the optimum sequence in its margins and the "crossover frequency" F_{ij} as its elements. RASC derives this matrix from the final order relation matrix of the ranking (the last S-matrix of the matrix-permutations). For the simple example used in the ranking part, this matrix-conversion can be illustrated, as follows:

		A	B	C				A	B	C
	A	X	4.5	5.5			A	X	$\frac{4.5}{6}$	$\frac{5.5}{7}$
S-matrix	B	1.5	X	5		F-matrix	B		X	$\frac{5}{5}$
	C	1.5	0	X			C			X

(A occurs above B in 4.5 sections)   (A occurs above B in 4.5 out of 6 sections)

RASC now converts the F-matrix into a Z-matrix, to obtain a set of direct estimates for the inter-event distances ($\Delta_{AB} = z_{AB}$)

$$F_{AB} = 4.5/6 \rightarrow P_{AB} = 4.5/6 = 75\% \rightarrow z_{AB} = 0.674$$
$$F_{AC} = 5.5/7 \rightarrow P_{AC} = 5.5/7 = 78\% \rightarrow z_{AC} = 0.788$$
$$F_{BC} = 5/5 \rightarrow P_{BC} = 5/5 = 100\% \rightarrow z_{BC} = \infty$$

- 15 -

This last value makes sense because B always occurs above C, which makes the statistical distance equal to ∞. A situation like this may very well arise when a large number of events over a long time-interval is used. Crossing over between relatively young and relatively old events never takes place of course.

Since ∞ is not a practical value, RASC replaces it with an arbitrary number q (or for P_{AB} = 0% by -q) and states that whenever z ≥ q or z ≤ -q, the values are to be ignored and consequently do not contribute to any distance estimates.

q can be set by the user. It has to be larger than 1 and smaller than 2. In most past uses of RASC q has been given a value of + 1.645, which corresponds to P_{AB} = 95% (or 5%).

It should also be noted that some scores in the original S-matrix were zeroed. They are the scores for which $r_{ij} < m_{c1}$, r_{ij} being how often events i and j occur together and m_{c1} being the second threshold parameter set for the matrix-permutations. These zeroed scores are given a z-value of 9.000 and are also ignored.

The Z-matrix for the example is given below. z_{AB} is the value for the direct estimate of the mean distance: Δ_{AB}

	A	B	C
A	X	0.674	0.788
B		X	1.645 (q)
C			X

(4) Unweighted indirect distance estimate ($\overline{\Delta_{AB}}$u)

The z-matrix contains the direct estimates for the statistical distances between events (Δ_{ij}). However, these direct estimates are considered insufficient, since i and j do not very often occur together in a section.

- 16 -

Moreover there is more information available, since other distance estimates can be used for the estimation of Δ_{ij}. When event C is used for the estimation of Δ_{AB} (which is then written as $\Delta_{AB.C}$ or $\overline{\Delta_{AB}}$), $\Delta_{AB.C} = \Delta_{AC} - \Delta_{BC}$:
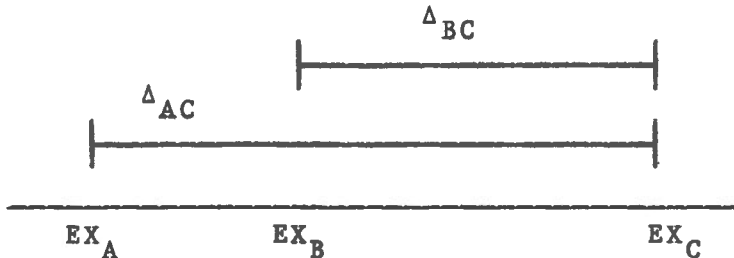
The distances Δ_{AC} and Δ_{BC} are calculated from their cross-over frequencies F_{AC} and F_{BC}. It follows that:

$$\Delta_{AB.C} = z_{AC} - z_{BC}$$

When N* events are used, the distance $\overline{\Delta_{AB}}$ is thus estimated by RASC in (N*-1) ways. The following estimate for $\overline{\Delta_{AB}}$ is computed:

$$\overline{\Delta_{AB}}^{u} = \frac{z_{AB} + (z_{AC} - z_{BC}) + (z_{AD} - z_{BD}) + \ldots\ldots}{N^* - 1}$$

Note: N* is not always equal to the total number of events (N) in the margins of the Z-matrix. Sometimes z-values have been set equal to 9.000 or q and are now ignored.

(5) <u>Weighted indirect distance-estimates</u> $(\overline{\Delta_{AB}}^{w})$.

The indirect distance-estimate is based on very different r_{ij}'s--how often event i and event j occur together in a section. Each z-value is based on its own r_{ij}. For this reason RASC computes a weight for the mean distance Δ_{AB}:

$$\Delta_{AB}^{w} = \frac{w_{AB} z_{AB} + w_{AB.C}(z_{AC} - z_{BC}) + w_{AB.D}(z_{AD} - z_{BD}) + \ldots\ldots}{w_{AB} + w_{AB.C} + w_{AB.D} + \ldots\ldots\ldots}$$

- 17 -

The values for w_{AB} and $w_{AB.K}$, as used by RASC in the above equation, are primarily determined by r_{AB}, r_{AK} and r_{BK}. (How often A occurs together with B and how often A and B occur together with all the other events used for the indirect estimation). The values w_{AB} and $w_{AB.K}$ are explained below.

When the weight w_{AB} is assigned in the above way to z_{AB}, which is a random variable, the value for w_{AB} is per definition conversely proportional to the variance of $f(z_{AB})$:

$$w_{AB} = \frac{1}{\sigma^2(z_{AB})}$$

To relate this value to r_{AB}, which is the primary factor in determining the weight, one has to derive the functional relationship of the variance $\sigma^2(z_{AB})$ with the variance $\sigma^2(p_{AB})$ of the frequency distribution (binomial) $f(p_{AB})$, which has a mean of P_{AB} and a variance of:

$$\sigma^2(P_{AB}) = \frac{P_{AB}(1-P_{AB})}{r_{AB}}$$

That there is such a relationship can be gleaned from Figure 3. For the exact mathematical derivation one is referred to Brauer Hudson and Agterberg (1982). The relationship can be written as follows:

$$\sigma^2(z_{AB}) = \frac{2\pi}{\exp(Z^2_{AB})} \cdot \sigma^2(P_{AB})$$

The value for w_{AB} is now the following:

$$w_{AB} = \frac{1}{\sigma^2(z_{AB})} = \frac{r_{AB} \cdot e^{-z_{AB}^2}}{P_{AB}(1-P_{AB}) \cdot 2\pi}$$

in which $\sigma^2(z_{AB})$ is the variance of $f(z_{AB})$,

P_{AB} equals the observed frequency of F_{AB},
r_{AB} indicates how often events A and B occur together, and
z_{AB} is looked up in the table for cumulative standard
normal distributions.

The other weight used by RASC, $w_{AB.K}$, has the following
value:

$$w_{AB.K} = \frac{1}{\sigma^2 (z_{BK}) + \sigma^2 (z_{AK})} = \frac{1}{\frac{1}{w_{BK}} + \frac{1}{w_{AK}}} = \frac{w_{BK} \cdot w_{AK}}{w_{BK} + w_{AK}}$$

(6) Standard deviations

For the (stratigraphic) interpretation of the distance-
values in the output it is useful to know the standard-
deviations. All computing for Δ_{AB}, the direct distance-
estimate, was based on the variance $2\sigma^2$ (standard deviation
$\sigma\sqrt{2}$), but the standard deviation of the weighted, indirect
estimate $\overline{\Delta_{AB}w}$ is the square root of the following variance:

$$\sigma^2 (\overline{\Delta_{AB}w}) = \frac{w_{AB}(z_{AB} - w_{AB})^2 + w_{AB.C}(z_{AC} - z_{BC} + w_{AB})^2 + \ldots}{w_{AB} + w_{AB.C} + \ldots}$$

(7) Preliminary solution of scaling

The set of distances between consecutive events, as determined by scaling, appears in the output of the RASC-program in a table like Table 2. Table 2 is the "weighted distance" table of the sample run. The table of "unweighted distances" is available in the output as an option (see "program inputs", Chapter IV B).

Table 2 lists the inter-event distances in the third column, and the cumulative distances in the fourth column. These cumulative distances use the first event of the scaled sequence as the origin.

The fifth column of Table 2 is called "sum diff z-values" and it contains the sum of the indirect weighted distance estimates for a certain event-pair: In the case of events A and B, it lists $\Sigma w_{AB \cdot K}(z_{AK})$. The column "weight" contains the w_{AB} and the column, "s.d." contains the standard deviations.

(8) New sequence

Some of the inter-event distances turn out to be negative (Table 2). This is the result of the (weighted) indirect estimates for scaling as opposed to the direct unweighted considerations in ranking. In order to obtain all-positive inter-event distances the events will have to be rearranged into a new sequence. This rearranging, or sorting of events, is done on the basis of their cumulative distances. For this purpose RASC takes arbitrarily the position of the first event as the origin of the sequence and calculates the cumulative distance of each event to this origin (Table 2). The rearranging results in a "new sequence".

The "new sequence" of events is given in the RASC output in the "events sorted" table (Table 3). (The appearance in the output of "events sorted" tables for unweighted distance estimates is optional. See "program inputs".)

The "events sorted" tables do not list the standard deviations for reasons that will be explained under the "final reordering option" (this chapter).

- 20 -

(9) Dendrograms

     With the use of the subroutine DENDRO (adapted from
Davis 1973, Program 7.8) RASC now constructs a dendrogram of
the (new) sequence of events (Table 4). Dendrograms of the
result of the unweighted distance-estimates are, together
with their tables, also available, but their appearance in
the output is optional.

     The purpose of the dendrogram is to give the user a
visual illustration of the inter-event distances. The
inter-event distances are indicated with horizontal lines;
when a line is long, a distance is great, when a line is
short, the distance is small.

     The vertical axis of the dendrogram contains the events
(numbers and names) in their established sequence and the
distance-value of each event to the next event. On the
horizontal axis, the inter-event distances are represented
on a scale, so that the horizontal lines can be given scaled
lengths. To be able to immediately compare the lengths of
the horizontal lines, vertical lines are drawn from the end-
point of each horizontal line down to where it meets another
horizontal line.

     This way the longest horizontal lines indicate (time)
gaps in the sequence (great inter-event distances), while
the shorter horizontal lines, in between, represent more or
less tightly clustered events. The shorter the lines, the
tighter the clustering.

     The dendrogram serves thus as an illustration of the
clustering of events along the (time) scale.

(10) Final reordering option

     Since the estimates for successive distances are not
statistically independent, the standard deviations for the
"new sequence" (Table 3) cannot simply be listed. For this
reason a final optional routine has been added to the
program: "final reordering".

     In the final reordering option RASC repeats all calcu-
lations for the distance-estimates, starting with a new z-
matrix that has the new sequence of events in its margins.
The purpose is to obtain a new set of standard deviations.

In places where the sequence has changed these new distance-estimates will turn out to be different from the first result of scaling. This is caused by the use of different z-values for the indirect distance estimates: When before $z_{CD}-z_{BD}$ was used, now $z_{CE}-z_{BE}$ might have to be used.

Some of the new distances might even be negative, in which case the entire procedure will have to be repeated again. RASC allows at most four complete reorderings, after which most of the distances between successive events are positive, and their corresponding standard deviations are known. The only negative distances still left are very small. The result of the final reordering is the final Scaled Optimum Sequence.

C. MARKER HORIZONS AND UNIQUE EVENTS

When marker horizons and unique events are identified as such on the appropriate cards or lines (see "Programs Inputs", Chapter IV), in addition to being entered like all other events, they will be treated in a different way than the other events. A marker horizon, such as a bentonite bed resulting from a volcanic ash-fall, is treated as a stratigraphic event with 0 variance with regard to its position. A unique event is an event that only occurs in one or two wells but yet constitutes important information to the user (as is the case with an index fossil). The unique event is fitted in the scaled optimum sequence after termination of scaling and reordering.

It is important to note that for a marker horizon to participate it has to occur in at least k_c wells, while a unique event is only treated as such when it occurs in fewer than k_c wells.

The marker horizons are identified in the dendrograms with a single asterisk and the unique events only appear in the last dendrogram (after the normality tests) and are identified with two asterisks.

(1) Marker horizons

In order for an event to be treated like a marker horizon, the following assumptions are made:

(1) A marker horizon is a stratigraphic event with 0 variance with regard to its position; and

(2) The values for the distance between a marker horizon M and a stratigraphic event A (d_{AM}) are normally distributed with a mean of Δ_{AM} and a variance of σ^2. With any two other events, B and C,

$$z_{AM} = \Delta_{AM}/\sigma \rightarrow \Delta_{AM} = \sigma \cdot z_{AM} = \sqrt{0.5} \cdot z_{AM} \text{ (Assumption 4 of}$$
scaling: $\sigma = \sqrt{0.5}$)

$$\Delta_{AM.C} = \Delta_{AC} - \Delta_{MC} = z_{AC} - \sqrt{0.5} \cdot z_{MC}$$

$$\Delta_{AB.M} = \Delta_{BM} - \Delta_{AM} = \sqrt{0.5} \cdot z_{BM} - \sqrt{0.5} \cdot z_{AM}$$

The values for the weight w_{AM} and the indirect weight $w_{AM.C}$ or $w_{AB.M}$ have been adjusted accordingly.

If a data base contains two marker horizons, they will theoretically never cross over. A situation in which they do cross over cannot be dealt with by RASC and should be avoided. The appropriate warning is given in the program.

(2) Unique events

A unique event U is fitted in the scaled optimum sequence by identifying its neighbours in one of the sections in which it occurs. When its neighbours are A and B, and events A and B have scaled optimum positions X_A and X_B (defined by their cumulative distance from the first event in the scaled optimum sequence), the position of U in the scaled optimum sequence (x_U) can be calculated from X_A and X_B (Fig. 4):

In one Section	Scaled Optimum Sequence
Etc. A U-S B Etc.	ORIGIN • • • • X_A X_S X_B

Figure 4.   A unique event U in one section compared
to the scaled optimum sequence positions
of its neighbours (X_A, X_S, X_B), where
S is any event coeval with U.

$$\overline{x_U} = \frac{X_A + X_S + X_B}{3} \quad \text{or when event S is missing.}$$

$$\overline{x_U} = \frac{X_A + X_B}{2}$$

In reality the situation might be found that A is
simultaneous with K and L in the section studied, and that B
is simultaneous with, say, event N (Fig. 5).

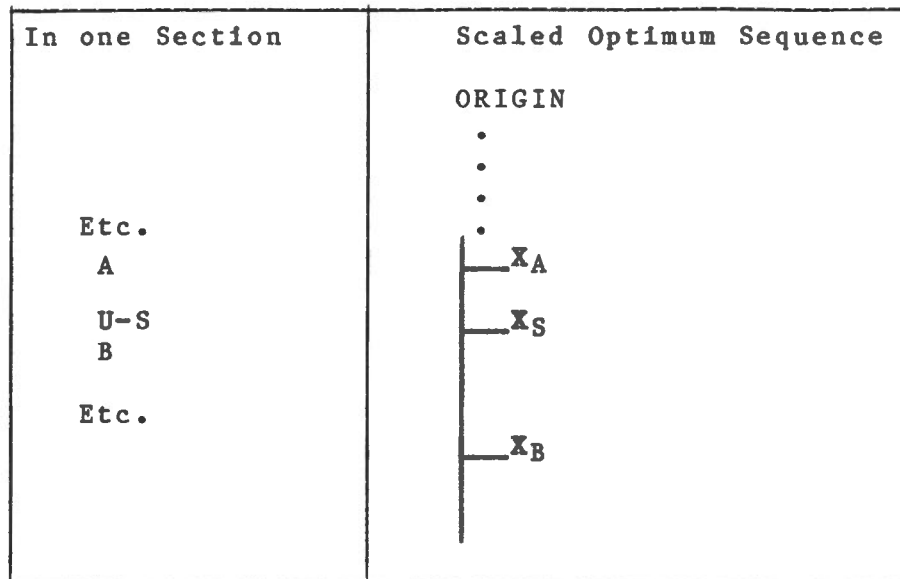| One Section | Scaled Optimum Sequence | | |
	Sequence	1st estimate	2nd estimate
	ORIGIN	ORIGIN	ORIGIN
	.	.	.
Etc.	.	.	.
	.	.	.
A-K-L	X_L	↑	
U-S	X_K		
B-N	X_A	1/2 R	
Etc.		↓	$\overline{x}$
	X_S	$\overline{x_U}$	
	X_B	↑	
		1/2 R	
	X_N	↓	
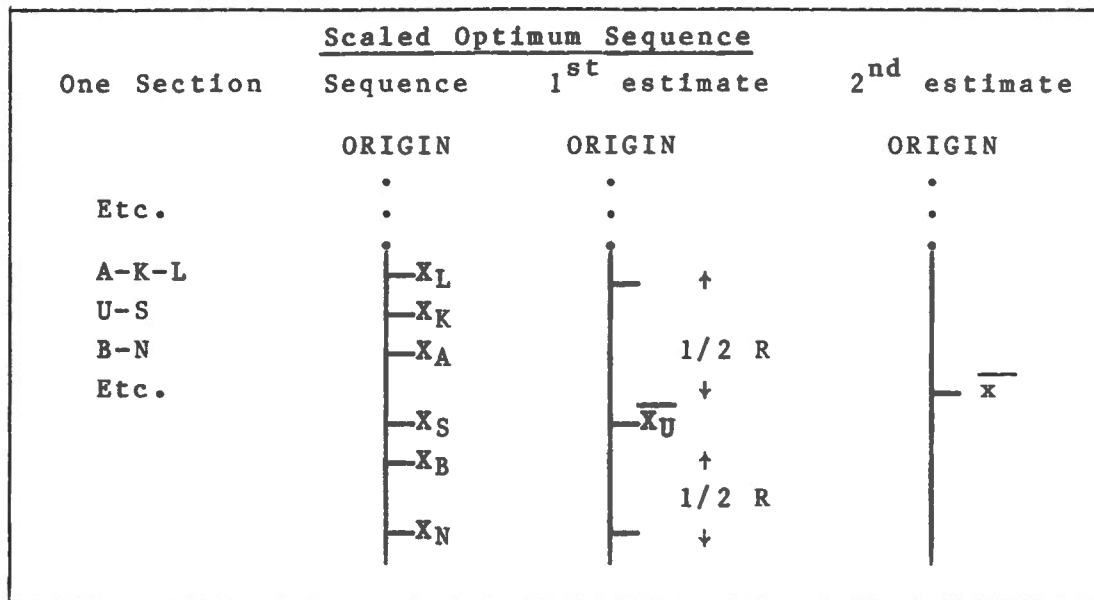
Figure 5.   A unique event U in one section compared to the scaled optimum sequence positions of its neighbours, to estimate a position for the unique event in the scaled optimum sequence ($\overline{x_U}$).

Instead of X_A, now an average of X_A, X_K, and X_L is used, and instead of X_B the average of X_B and X_N.   A preliminary estimate for x_U is then obtained by averaging the average X_A and the average X_B (Fig. 5).

The ultimate estimate for x_U is obtained by introducing a range R for the preliminary x_U : $x_U \pm 0.5R$.   This range includes all events encountered within the vicinity of x_U with a probability greater than 5%.   Since x_U is normally distributed with a variance of σ^2, and $\sigma^2 = 0.5$ (assumptions 1 and 4) → R = (1.965.σ =) 1.38.   In the example X_A, X_K and X_B fall within this range, and the ultimate estimate for x_U is obtained by averaging ($x_U + 0.5R$) with X_A and X_K, and by averaging ($x_U - 0.5R$) with X_B, and by adding these averages to X_S (or the average of X_S and scaled positions of other events that were simultaneous with S and U in the section studied) and dividing the sum through 3.

If the unique event occurs in more than one section, the preceding calcuation is carried out for each section separately and the resulting values for $\overline{x_U}$ are averaged.

D. NORMALITY TESTING

Once the scaled optimum sequence of events and the cor-
responding set of distances between consecutive events has
been obtained, RASC provides four options for comparing the
event-positions in the individual sections to the positions
in the scaled optimum sequence. These options are:

(1) Occurrence table, in which the events in the
     individual sections are recorded in their scaled
     optimum sequence.

(2) Step model table, in which each event in each
     section receives a penalty score indicating how
     often it is out of place in that section.

(3) Scattergrams, consisting of a series of bivariate
     plots in which the order of events in each section
     is plotted against the order of events in the
     optimum sequence. These bivariate plots can be
     produced for both the ranked optimum sequence and
     the scaled optimum sequence.

(4) Normality tests; (a) a series of statistical tests
     comparing the position of each event in each
     section to its position in the scaled optimum
     sequence; (b) the "difference-values" found in
     this way for each event in each section (except
     for the first and last event of each section) are
     then together tested against an assumed normal
     distribution.

The purpose of these comparisons is to detect events
that occupy anomalous positions in an individual section,
which might be the result of geological reworking, sample
contamination or misidentification. The comparisons can
also lead to conclusions about homogeneity of the region
studied.

(1) Occurrence Table

The occurrence table is constructed with the scaled optimum sequence of the events in its vertical direction, and the individual sections in the horizontal direction (Table 5). Events that occur in a certain section are indicated with an x. In this way, the events that occur in a certain section appear in that section according to their optimum sequence. This table is useful for a quick first glance interpretation of the result.

(2) Step Model

Each event in a section receives a penalty point each time it is out of place in relation to the other events in that same section, when that section is compared to the scaled optimum sequence. This way each event accumulates, in each section, a "penalty score" (Fig. 6). A high penalty score indicates that an event is likely to be out of place in that section.

Petrel A-62	Scaled Optimum Sequence	Penalty Score
4	4	0.0
14	13	2.0
15	14	1.0
13	15	1.0
72	72	0.0
27	27	0.0
131	131	0.0

Figure 6. Step model comparison of the Petrel A-62 well with the scaled optimum sequence, both of the sample run.

The step model has the following draw-back: Events which belong to clusters of events, with many internal inconsistencies, are likely to accumulate high scores, even if they are in their normal places. The step model does not provide a way to distinguish between anomalous events and clustering events.

(3)   Scattergrams

RASC scattergrams plot the order of events in the optimum sequence along the vertical axis and the order of events in an individual well across the top horizontal axis. Each event in the individual well is plotted against its position in the optimum sequence. Co-eval events in a section are stacked vertically in the same position. An example of a scattergram is given in Table 6. Scattergrams provide a quick, visual indication of the degree of deviation of the positions of specific events in each section from their respective positions in the optimum sequence.

(4)   Normality Tests

(a)   Statistical Tests for event-positions in
       individual sections

The difficulty in giving a definite value to the amount an event-position is out of place in a certain section, is that the positions in individual sections do not have values or distance values as their counterparts in the scaled optimum sequence do.

Therefore, in order to test a certain event-position in a certain section, the event's neighbours in that section are identified. All three of these events have a position in the scaled optimum sequence, and a value for their cumulative distance from the origin (the first event), of the scaled optimum sequence. These three position-values are compared, by calculating a "second order difference value". When the position values are very different (when the three events should not be neighbours) the "second order difference value" is very much above or below 0. When the three position-values are close, the "second order difference value" approaches 0, and the events "belong together".

- 28 -

Scaled Optimum Sequence		In a Section			
		sequence	cum. dist.	1st order diff.	2nd order diff.
H	-0.0	H			
J	-0.6	M			
K	-1.0	-L	1.7		
N	-1.5	-N	1.5	-0.2	
L	-1.7	-J	0.6	-0.9	-0.7
M	-1.8	0			
0	-2.0				

Figure 7. Example of calculation of 2nd order difference
value for event N in a certain section. Note:
hyphens in front of events in the section indi-
cate that those events occurred simultaneous in
that section.

The second order difference value is calculated as fol-
lows (Fig. 7): Event N has, in a certain section, events J
and L for its neighbours. The "scaled optimum sequence
position" of the "cumulative distance", of N (X_N) is sub-
tracted from the cumulative distance of J(X_j), and the
cumulative distance of L (X_L) is subtracted from X_N, and
these differences are subtracted once more.

(1) 2nd order difference $= (X_J - X_N) - (X_N - X_L)$

$$= (0.6 - 1.5) - (1.5 - 1.7) = -0.7$$

Since the event-positions in the individual sections
are expected to be slightly different (a gradual increase
from the origin has to be taken into account), the first
order difference is corrected by a small amount when non-

simultaneous events are compared.[1] Simultaneous events don't need this correction.

Algebraic manipulation of (1) shows that the second-order difference calculation can be given as the minus of the difference between twice the distance of an event on the one hand and the sum of the distance of its two neighbouring events on the other. Rewriting (1) we have:

(2) 2nd order diff. $= -[2 X_N - (X_J + X_L)]$

An assumption during scaling was that each event position can be regarded as a realization of an independent normal random variable with variance equal to σ^2. If we regard successive differences, similarly, as realizations of independent normal random variables with variances equal to $2\sigma^2$, the variance of the second-order differences, σ^2_2, would equal $6 \sigma^2$. However, since successive distance estimates have become auto-correlated due to various manipulations to which the data were subjected during ranking and scaling, an auto correlation coefficient, ρ, must be estimated. This estimate is required in determining the number of independent values involved in the normal distribution curve. (See Section 4b)

RASC estimates the normal distribution curve by fitting a doubly-truncated normal distribution to the central 60% of the observed second order differences. The second order differences are ordered, from all sections, from the smallest value to the largest one. The standard deviation, σ_2 of the central 60% of the ordered values is determined and assumed to represent a truncated normal distribution. From statistical tables, the standard deviation of a truncated normal distribution, σ_2, is approximately 0.463 the standard deviation of the corresponding full normal distribution, $\hat{\sigma}_2$ (if 20% is truncated from each tail). Division of σ_2 by 0.463 yields the estimate $\hat{\sigma}_2$. It is this standard deviation that is used by RASC to identify anomalous events, in each section at the 95% and 99% confidence levels. The 95% and 99% probability units of this estimated normal distribution are given from statistical tables as $\pm 1.960 \hat{\sigma}_2$ and $\pm 2.576 \hat{\sigma}_2$ respectively.

---

[1] The small amount was set equal to the difference between highest and lowest position values in the observed sequence divided by the total number of times an event precedes another event without being simultaneous with it.

In the RASC normality test output, a table like Figure 7 appears for each well (e.g. Table 7). When the second-order differences in the last column of these tables exceed the 95% or 99% probability limits given above, the corresponding event positions are considered to be out of place or anomalous and are flagged by a single or double asterisk respectively. The reason for restricting the estimation of the normal distribution curve to the central 60% of the second-order difference values, is to avoid the contribution of anomalous values, which, if present, are more likely to occur in the tails of the distribution.

It should be noted that no 2nd order difference values can be determined for the first and the last events of each section.

In practice an anomalous event often turns out to have another anomalous event for its neighbour in a section. This is an artifact of the method. The test cannot distinguish between an anomalous event and its neighbour.

It should also be noted that if there are large hiatuses in an individual section, the events on both sides of such a hiatus will give rise to very high (or very low) 2nd order differences which do not necessarily mean that the events positions are anomalous in that section.

(b) The distribution of all 2nd order difference values of the data base: Chi-Square Test

The second part of the normality test attempts to assess how well the assumption and estimation of a normal distribution for event positions fits the actual total data set. This is accomplished by comparing the frequency of observed second-order differences with the frequency of expected second-order differences (on the assumption of normal distribution) within a set of equal probability classes.

For this purpose the area under the expected normal distribution-curve is divided into 10 equal parts (Fig. 8). The 10 areas represent 10 classes with identical occurrence frequencies, each indicating how often a specific second order difference value is expected to occur. The ten areas are called 10 classes of equal probability and the limits of each class are the second order difference values that belong to a cumulative frequency of respectively 0.1, 0.2, 0.3, etc. (Fig. 8). They are:
Class 1/ $\hat{\sigma}_2 \cdot z_{0.1}$ / Class 2/ $\hat{\sigma}_2 \cdot z_{0.2}$ / Class 3/ $\hat{\sigma}_2 \cdot z_{0.3}$ ... etc.
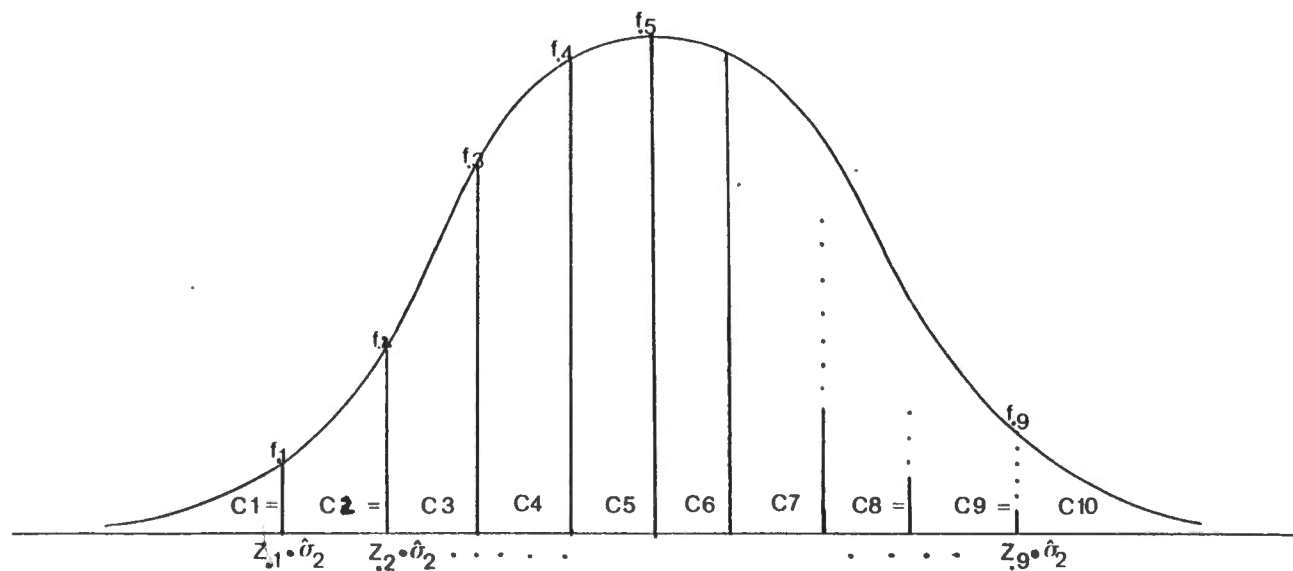
Figure 8. Normal distribution of 2nd order diff. values divided into classes of equal probability.

Each of these classes is <u>expected</u> to contain one tenth of all second order differences of the data-base. The "<u>expected occurrence</u>" of each class is thus 1/10 of the total number of second order differences.

By placing all <u>observed</u> second order differences in their appropriate classes one can test if they are indeed normally distributed. For this purpose each second order difference with a value within the limits of a certain class is placed in that class. The number of second order differences thus placed in a class constitutes the "<u>observed occurrence</u>" of that class.

Table 8 is an example of output, from the second part of the normality test, with an approximate chi-square test for goodness of fit to a normal distribution. In part 4a it was pointed out that the successive distance estimates have become auto-correlated from the manipulation performed during ranking and scaling. A measure of that auto-correlation will also affect the second-order differences calculated by RASC. Since chi-square is defined as the sum of squares of <u>independent</u>, normally distributed variables, the auto-correlated data must be related to an equivalent set of independent values before the chi-square test for normal distribution can be employed. From Agterberg, F. (1974, p. 302) the relationship between n auto-correlated values and n^1 stochastically independent values can generally be assumed to satisfy the following:

$$\frac{1}{n_1} = \frac{1}{n} + 2\,\hat{\rho}\,\left[\frac{n}{(1-\hat{\rho})} - \frac{1}{(1-\hat{\rho})^2}\right]\,/n^2$$

where $\hat{\rho}$, rho, represents the auto-correlation coefficient applied to successive distances. Both n^1, the number of equivalent independent values and $\hat{\rho}$, rho, are part of the RASC program's output and appear under the second order differences statistics heading just before the output for the individual normality tests. An example of these statistics is given below

n = 211   ave = - .06769   SD = 1.47050   Ratio = .85
rho ($\hat{\rho}$) = .222 n^1 (equivalent number of values) = 135

The value, <u>ratio</u>, represents the approximated second-order difference standard deviation (S.D.), $\hat{\sigma}_2$, divided by the theoretic second-order difference S.D., $\sqrt{6\sigma^2}$, where $\sigma^2 = 0.5$ from assumption 4 of scaling.

- 33 -

In the chi-square test for goodness of fit, expected frequencies, E_i, of stochastically independent data in 10 classes are related to the corresponding observed frequencies, O_i, by

$$\chi^2 = \sum_{i=1}^{10} (E_i - O_i)^2 \, / \, E_i$$

with seven degrees of freedom since two statistics, population mean and standard deviation have been estimated. The RASC program provides an approximate chi-square test for auto-correlated data by multiplying the right hand side by n^1/n. The approximated chi-square is defined, then, as

$$\hat{\chi}^2 = \sum_{i=1}^{10} \left[\frac{(E_i - O_i)^2}{E_i} \right] n^1/n$$

As in 4a, deviations that are statistically significant at the 95% and 99% confidence levels are flagged by a single or double asterisk respectively.

I M P L E M E N T A T A T I O N   O F   T H E   R A S C   P R O G R A M

A.   PREPARATION OF THE DATA

To implement the RASC-program for a set of raw data, the following steps have to be taken:

(1) Listing of the raw data. The list should contain the section-names, and for each section a sequence of samples with a list of events for each sample (Table 9).

(2) Production of a dictionary. Each event of the list has to be assigned a unique number. The dictionary has to be key punched on Fortran-cards or entered from a terminal keyboard.

(3) Coding of the observed event-sequences in the individual sections, using the numbers as assigned in the dictionary.

(4) Unique events and marker horizons have to be identified.

(5) Threshold parameters k_c, m_{c1} and m_{c2} have to be established.

(1) Raw Data

Until now RASC has only been used for the ranking of
bio-stratigraphic events, but it could be used for any aver-
age ranking of elements along a relative scale. Example:
RASC can list the average order in which the 12 yachts of
the Bedford Basin Yacht Club have passed the finish in the
20 regattas they participated in last summer. The data
should have a relative order, observed in different trials
(sections etc.).

In bio-stratigraphy one could presumably make a list
of taxa, one following another, but since taxa show overlap
in time, it is better to list events that happen to the
taxa. At one time a certain taxon, say the planktonic
taxon Globotruncana contusa, first appears on the fossil
record. "First appearance of Glob. cont." can be listed as
an event and given a number, say 1. The same taxon can dis-
appear from the fossil record higher up in the section.
"Disappearance" or "Top of Glob. cont." can be listed as
another event and given a number, say 2.

The taxon Globotruncana stuarti can thus be involved
in events 3 and 4, and so on.

When RASC is to be used for a certain data-base, it
should first be decided which of the events, in which each
taxon is involved, can be identified. In the data of the
sample run, the above taxa G. contusa and G. stuarti are
both listed with their "last" or "top" occurrence. Since
the data base for the sample run is based on well-cutting-
samples, the event "first appearance" could not be identi-
fied. When cores are used, or outcrops are sampled, many
events can be identified for each taxon (i.e. "top", "sub-
top", "subbottom", "bottom" etc.).

The user of RASC has to begin by listing all events,
in the observed order, by their names (i.e. top of
Globotruncana area) for each section or well. An example
of such a list is shown in Table 9. The final list shows
the observed order of all events in all sections. It
should be noted that this RASC program is limited to 25
stratigraphic sections.

(2)   The Dictionary

As an intermediary stage between the user's data and
analysis by the RASC program, the purpose of the dictionary
is to number each event in a manner compatible with the RASC
program. Each distinct event is listed and assigned a
unique number. Since the RASC program internally numbers
each event in the sequence in which it is entered, the dic-
tionary must list its assigned numbers consecutively and
continuously and it must be entered into the computer in
that order. How you list your events originally in the dic-
tionary is an arbitrary choice in that there is no require-
ment for a particular event to receive a specific number.
The only requirement is that list then be numbered consecu-
tively. [Note: RASC is restricted to process only 280 num-
bers or less. One's database should therefore not exceed
280 events.]

In practice, when there are large data bases, the dic-
tionary may require extensive editing. The amount of edit-
ing can be minimized by alphabetically listing the events
before numbering them consecutively. Errors may arise in
one of two ways and are handled as follows: (1) An event
may be mistakenly assigned two numbers. You should arbi-
trarily assign one of the numbers to a 'dummy' event, there-
by preserving the established numeric ordering. (2) An
event may be discovered to have been omitted after the dic-
tionary has been numbered. Simply place the 'discovered'
event at the end of the dictionary list (out of alphabetical
order) and assign it the next consecutive number. Again it
should be pointed out that it is the numeric order of the
dictionary list that is important, not the event order.

For the punching (or entry) of the dictionary, indi-
vidual cards (or lines) are used for each event. Each card
(or line) is allowed forty characters starting in column 1
in 10A4 format. Since RASC contains a subroutine for an
alphabetical listing of events, alphabetical characters must
precede numerical characters. A last, extra entry in the
dictionary record must be the word 'last' placed in the
first four columns of a final card (or line). This marks
the end of the dictionary record.

A separate file must be created for the dictionary.
The dictionary events are then read into the RASC program
from this separate file, called 'Tape 99' (see "program
inputs").

(3) List of Coded Sequences

Once all events have been assigned unique numbers in
the dictionary, the sequence of events in each section is
numerically coded, identifying the events by their assigned
numbers and preserving their observed sequences. Events
which appear in the same sample of a section (or well) are
considered to occur simultaneously. Simultaneous occur-
rences are identified as such by a dash or hyphen sign
between them. For example, in the Petrel A-62 well of the
sample run the following order of events has been observed
(Table 9):

    4   14   15   13   72   27   33-170   131   169

The events are ordered from left to right, representing
either a stratigraphically upwards or downwards sequence.
Which type of sequence is represented is decided in the
ordering of the original data. A hyphen before an event
indicates that that event occurs simultaneously with the
event immediately preceding it in the sequence. In this
example, events 33 and 170 occur simultaneously.

With the completion of the list of coded sequences, the
data can now be key-punched and read into the RASC program
either by cards or disc file. The format for this procedure
is contained in the listing of the RASC computer program.

In keypunching the list of coded sequences, the first
punched card contains the name of the section (well) only,
in 5A4 format. Twenty columns are available for the title
name. The remaining six cards contain the sequence numbers.
The (20I4) format indicates that the 80 columns of the
fortran card or record are divided into 20 units of four
columns. Each unit is occupied by an event number, with a
maximum of three digits, with or without a hyphen. Empty
units or fields, are ignored. The numbers of the coded
sequence are punched in their respective order from left to
right. There are no column spaces between units, conse-
quently each card will contain twenty numbers, (if there are
no empty fields) except for the last number card which may
contain less. For example, in key punching the Petrel A-62

- 38 -

well of the sample run, card one gets the name "Petrel A-62"
only. Card two is punched as illustrated in Figure 9.
Since Petrel A-62 has only 10 events, punching ends at 169,
and the rest of the card is left blank, as is the case with
the remaining 5 cards for that well. To indicate the end of
a well sequence, the last entry or record should be -999.
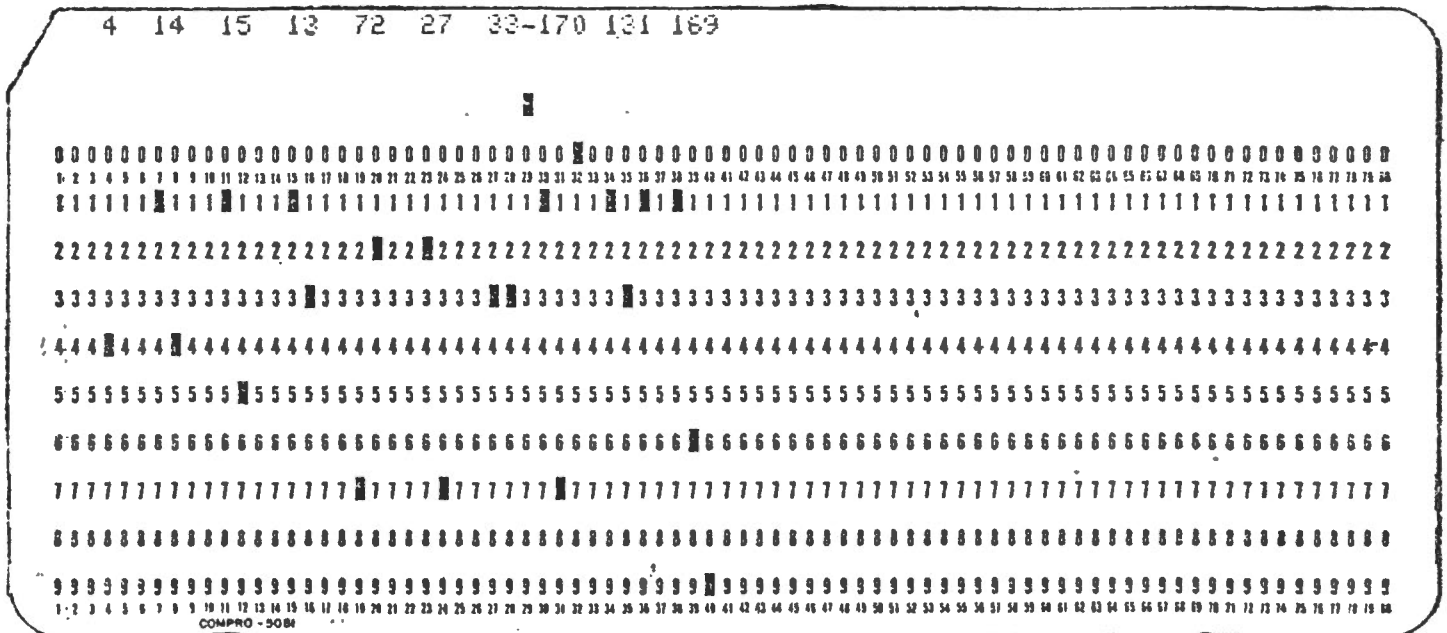
Figure 9. A fortran card containing the coded sequence of
the Petrel A-62 well of the sample run.

(4) Marker Horizons and Unique Events

If there are events in the dictionary that are to be
treated as Marker Horizons or Unique Events, those events
first have to be treated as ordinary events (re: the
dictionary and the list of coded sequences), but then they
should also be identified on separate, designated cards or
lines (see program inputs).

- 39 -

(5)   Threshold Parameters

Many of the listed events occur in only one or very few sections -- they are not "common". Therefore a threshold parameter k_c has to be set. "k_c" will allow only those events to participate that occur in at least k_c wells. For the matrix permutations involved in ranking and the distance analysis in scaling, two other threshold parameters are needed m_{c1} and m_{c2}. m_{c1}, or CRIT1 which should be set at less than or equal to k_c, allows only those pairs of events to participate in the determination of the ranked optimum sequence that occur in at least m_{c1} wells or sections.

m_{c2}, or CRIT2, defines a minimum sample size of event pairs that must be met in calculating the relative frequency or cross-over ratios used for the distance estimations in scaling. m_{c2} restricts the sample size to only those pairs of events that occur together in at least m_{c2} wells. m_{c2} must be set greater than or equal to m_{c1}^2 during a RASC run.

It should be kept in mind that, although the threshold parameter is a number of wells, its use limits the number of events and not the number of wells.

For the first run with RASC the threshold parameters have to be chosen arbitrarily. In the RASC-output, under the heading "preprocessing initiated", tables appear that allow for a more rational choosing of values for k_c, m_{c1} and m_{c2}. All events are listed by their dictionary numbers versus the number of sections in which each of the events occurs. This table is followed by a shorter version (Table 10) in which the second row indicates how many events occur in a certain number of wells and the third row how many events occur in at least a certain number of wells. Note: The information in both rows is the same but the presentation differs.

In the sample run were 76 participating events (an
original dictionary of 184 events was truncated for the
purpose of the sample run), which makes for 76 events that
occur in at least 1 well (third row of Table 10), although
there are only 10 events that occur only in 1 well (2nd
row, Table 10). There are 9 events that occur in 4 wells
(2nd row, Table 10), but there are 40 events that occur
in four or more wells.

For the sample run k_c was set at 5, which means that
30 events participated in ranking and scaling. m_c has to
be lower, it was set at 2 and m_{c2} was set at 3.

After the user has entered k_c, m_{c1} and m_{c2} into
the input (see "program inputs"), RASC automatically edits
the list of coded sequences to exclude the appropriate
events.

B. PROGRAM INPUTS

(1) Run Parameters

The first card or line of input into the RASC program contains the list of "Run Parameters". These are as follows: NS, IOCR, INIQ, TOL, ITER, CRIT1, AAA, CRIT2. A complete description of these parameters with their column locations and format is printed in the beginning of the listing of the RASC program. NS is simply the number of sequences or wells in the data (Note: In the present form of RASC, NS cannot be higher than 25). The codes IOCR, CRIT1 and CRIT2 designate the k_c and the m_{c1} and m_{c2} threshold parameters respectively. If either unique events or marker horizons are included in the data set, the parameter INIQ is set to 1. Otherwise it is set to 0. TOL stands for tolerance. When scores (s_{ij}) in the upper triangle of the cumulative order matrix are compared with their counterparts in the lower triangle (s_{ji}), s_{ij} may be lower than s_{ji} by as much as TOL. AAA designates the value for q (see scaling), and is often set at 1.645. ITER is the maximum number of matrix permutations that can be performed during ranking (See Chapter IIIA).

(2) Processing Control

Card 2 or line 2 of input gives a list of ten processing options or "control parameters". They are the following: ITAPE, IOMAT, ISRT, IALPHA, ITAB1, ISCORE, ICOMP, ISKIP, IFIN, INOSC, INEG and ISCAT.

ITAPE indicates whether the input are to be read from cards or disc file. When cards are used ITAPE is set at 0. When discs are used ITAPE is set at 1.

IOMAT controls the printing or suppressing of certain output data. If IOMAT = 1, printing of order and frequency matrices as well as the intermediate tables takes place. If IOMAT = 0 they are suppressed.

ISRT gives the user the option to use the presorting process, to shorten computer time. If ISRT = 1 the presorting option is used. If ISRT = 0 ranking is carried out by means of matrix permutations only.

IALPHA allows the RASC user to choose between a short version of RASC (ranking only) and a complete version (ranking as well as scaling and normality testing). If IALPHA = 0 RASC will terminate after ranking. If IALPHA = 1 scaling analysis will be carried out. If IALPHA = 2 the individual matrix permutations will be printed on hard copy, before scaling is terminated.

ITAB1 = 1 means that the occurrence table which precedes the normality test, will be printed.

ISCORE = 1 means that the step model will be performed.

ICOMP = 1 means that the normality test will be carried out.

ISKIP allows the user to skip the matrix permutations, which means that when ISKIP = 1 the ranking solution will be based on presorting only, which is not sufficient. It is therefore suggested to set ISKIP at 0.

IFIN stands for the final reordering option. When IFIN = 1 four complete reorderings will be carried out, if necessary, after scaling.

INOSC controls the printing of scaling output. (Note: It controls the printing of certain tables, not the performing of scaling itself.) If INOSC = 1 the unweighted distance output (tables and dendrograms) is suppressed. If INOSC = 2 weighted as well as unweighted distance output is printed.

INEG = Since distance estimates based on few pairs of z values are relatively imprecise, INEG provides the user with the option of suppressing large distance estimates based on few z values. By setting INEG = 1, if the number of values is less than m_{c2}, the distance calculation will be suppressed. If INEG = 0 all calculations will be performed.

ISCAT = 1 means the scattergrams or bivariate plots will be produced.

(3) Unique Events

Card or line 3 identifies the unique events of the data set. If INIQ = 1 on card 1, up to twenty unique events can be punched on card 4 in 20I4 format. If there are no unique events this card or line is left blank. If unique events are included, ICOMP on card or line 2 must be equal to 1. Note: If unique events occur in more than k_c (IOCR) wells they will not be treated as such.

(4) Marker Horizons

Card or line 4 identifies the marker horizons of the data-set. When INIQ = 1 on line or card 1, up to 20 marker horizons can be entered in 20I4 format on card or line 5. If there are no marker horizons this card or line is left blank. Note: The marker horizons have to occur in more than k_c (IOCR) wells to participate in ranking and scaling.

(5) Data Set

The data set contains the list of coded sequences in the format described earlier. This data can be read either from an external tape or file, referred to as TAPE10 by RASC, or from cards positioned immediately after card number 4. If ITAPE on card 2 equals 1 this data set will be read off external TAPE10. If ITAPE = 0, cards positioned after card 4 will be used to read in this data.

(6) Dictionary of Events

The format of the dictionary has been described previously. The dictionary of events must be placed on a separate file from RASC. It is read into RASC from this separate file, designated as TAPE99 by the RASC program.

- 44 -

To effectively illustrate the procedures involved in
RASC, an analysis of actual data was run through the RASC
computer program. Tables in the previous chapters were all
output from this sample run. The present chapter will
describe the input and give a short discussion of the
output.

A. DATA

The stratigraphic information used in the sample run
is based on Gradstein et al. (1984) and Williamson (in
press)). The data consists of 76 highest stratigraphic
occurrences of foraminifera of Late Cretaceous age, as found
in 20 exploratory wells in the Newfoundland Basin and on the
Grand Banks, N.W. Atlantic. For the purpose of the sample
run, the data of an original list of 184 Jurassic-Cretaceous
events in 25 wells have been truncated.

In a few cases a distinction has been made between the
highest occurrence, or top, of a taxon (as done for all the
other taxa), and the peak occurrence (subtop with a large
number of specimens) of that same taxon. These two events
of the same taxon have been given separate dictionary
numbers.

B. INPUT

A dictionary (Table 11) was prepared from the raw data
and after truncation of most of the wells and the exclusion
of some 7 more wells, a list of coded wells (sequences) was
constructed (Table 12). An example of the raw data for one
of the wells has been given in Table 9 (Chapter IV). In
this table the dictionary numbers were listed also, which
has to be done prior to coding of the well. After thus
preparing the data, the dictionary and the list of coded
sequences was key-punched as described earlier.

The run parameters for the sample run were chosen as follows:

NS   =   20   (20 wells)
IOCR   =   5   (events had to occur in at least 5 wells)
INIQ   =   1   (we had unique events or marker horizon)
CRIT 1   =   2   (only scores of events that occur together in at least 2 wells were considered for the matrix permutations)
TOL   = 0.0   (s_{ij} in the upper triangle must be higher than s_{ij} in the lower triangle)
AAA   = 1.645   (when i always occurs above or always occurs below j, the "distance" between i and j equals ± 1.645)
CRIT 2   = 3.0   (distance estimates were based on cross-over frequencies of pairs of events that occur together in at least 3 wells).
ITER   = 35000   (we allowed for 35000 matrix permutations to take place during ranking)
ITAPE   =   1   (we had our dictionary and our list of coded sequences on tape).
IOMAT   =   1   (we wanted to have all matrices and tables printed).
ISRT   =   1   (we used the presorting option)
IALPHA   =   1   (we used ranking as well as scaling)
ITAB1   =   1   (we had the occurrence table made and printed)
ISCORE   =   1   (we had the step model table printed out)
ICOMP   =   1   (we used the normality tests)
ISKIP   =   0   (we did not skip the matrix-permutations)
IFIN   =   1   (four complete reorderings are carried out, if necessary, following scaling. As can be seen in Table 2 (scaling) there were 7 negative inter-event distances, which makes for a quite different "new" sequence and list of inter-event distances (Table 3). This means that we cannot really take the standard deviations of Table 2 into consideration (see "Discussion of the Output")
INOSC   =   2   (we had all scaling output printed)
INEG   =   0   (no distance estimates were suppressed)
ISCAT   =   1   (bivariate plots produced)

The information in Tables 11 and 12 is also available on tape and can be used for a sample run by future users.

C. OUTPUT

The output of the RASC program contains the following eight sections: Input, Preprocessing, Presort Option, Ranking solution, Scaling Analysis, Final Reordering, Normality Tests and Final Dendrogram.

(1) Input. This (untitled) section contains input parameters and a print-out of the list of coded sequences (Table 12), which is called "original sequence data".

(2) Preprocessing. This section contains two tables of event occurrences, which allow for a rational choosing of values for k_c (IOCR), m_{c1} (CRIT1) and m_{c2} (CRIT2) (see threshold parameters", Chapter IV). It also contains a new set of coded sequences which are edited to exclude the events that do not occur in k_c wells.

(3) Presort Option. The "code followed by score" - table gives the presort scores (i.e. the Presort Rank Numbers) as calculated from the cumulative order matrix (see Ranking, Chapter IIIA). After the Presort Rank Numbers ("scores") have been grouped in descending order, the original sequence data is edited according to this new "Presort Sequence" and printed.

(4) Ranking Solution. This section contains the cumulative order matrixes before and after the matrix permutations (if IALPHA had been 2 it would have contained all intermediate matrices too). In addition it contains a printout of all cycles that occurred during the matrix permutations and which of the scores were zeroed because of the cycles. This part of the print-out begins with the statement that x events were zeroed, but they are the scores that were zeroed on the basis of m_{c1}.

Although the "optimum sequence obtained via rank-ing" follows directly after the cycles, the table of the Optimum Sequence with the ranges and the names of the events (as in Table 1) only appears after the printout of the dictionary and of the alphabetical list of events.

(5)   Scaling Analysis. This section contains matrices
        which are called "upper triangle of normal z-
        values", and which are a combination of a Z-matrix
        and an F-matrix (see Scaling, Chapter IIIB). The
        section further contains "distance" - tables and
        dendrograms, first for unweighted estimations and
        then for weighted estimations. The weighted out-
        put (tables and dendrograms) are explained in
        Chapter IIIB (scaling) and will be discussed
        below.

(6)   Final Reordering. Only the tables and dendrograms
        (unweighted and weighted) for the result of the
        last reordering are printed out. However, it says
        how many reorderings were required ("solution
        after x iterations", never more than 5, the first
        plus the four complete reorderings that RASC
        allows for [Chapter IIIB, final reordering]).

(7)   Normality Tests. This output (Occurrence table,
        step model table, for each well, a normality test
        table and scattergram or bivariate plot and the
        chi-square test) is explained in the Normality
        Tests section of Chapter III.

(8)   Final Dendrogram. This final dendrogram comes
        after the normality tests and is the only dendro-
        gram that contains the unique events, if they were
        input into the program.

Discussion of the output

     A detailed discussion of the biostratigraphic informa-
tion contained in the data and analyzed in the Sample Run
is beyond the scope of this manual. We limit ourselves to
a brief account of the meaning of some of the results.

     The first major result of RASC, the Optimum Sequence
is shown as Table 1 of this manual. It contains 31 events,
most of which are part of standard late Cretaceous Zona-
tions. The top few events have a long range, a result of
the lack of data above position zero. The Optimum Sequence
can be used for point correlation of well sections and can
serve to subjectively study stratigraphic normality of indi-
vidual wells, using bivariate Shaw-type diagrams. For such
a diagram the Optimum Sequence is plotted along the y-axis,

and the individual well sequence along the x-axis. The
resulting scatter diagram gives a visual expression of
regional biostratigraphic coherence. The Step Model and
Normality Test output of RASC also evaluate this feature.

A perusal of this Step Model and the Normality Test
output of the Sample Run shows that not many wells have
"badly positioned" events. Table 8 shows the data set to be
quite "normal", or reasonable.

The dendrogram of Table 4 contains four or five (stra-
tigraphically meaningful) clusters of Cenomanian through
Maestrichtian age. We will have a closer look at the two
apparently greater gaps in the scaled sequence: The
"distance" between events 20, 108 and 23 and the "distance"
between events 31 and 27.

Event 108 (top of Gavelinella minima) has in the Opti-
mum Sequence, after ranking, an exceptionally long Range
(Table 1). Its position in relation to events 26, 56, 20,
23 and 55, could not be determined, either because 108
occurs simultaneous with these events or because 108 did not
occur in enough sections together with them. After ranking
one can only say that 108 can be placed somewhere between
event 15, on position 12 (top of Globotruncana marginata)
and event 109, on position 19 (top of Coarse Agglutinated
spp.). After scaling event 108 is placed between the top
occurrence of Globotruncana coronata and top of Globotuncana
angusticarinata.

The gap between events 31 and 27 is better defined.
Event 31 (top of Praeglobotruncana turbinata) has a well-
defined position after ranking, between its predecessor (top
of Praeglob. stephani) and its successor (top of Rotalipora
cushmani). The standard deviation of the great "distance"
between top of Praeglob. turbinata and top of Rotalipora
cushmani is also relatively small ($.9363 \pm .1438$). It can
be concluded that the RASC analysis of the sample run has
revealed a gap in the sequence between the "top of"
Rotalipora cushmani and the "top of" Praeglobotruncana
turbinata. This gap may in fact be a regional hiatus.

This discussion is presented here to show that infor-
mation from different output tables of RASC should be con-
sidered before drawing conclusions about the final dendro-
gram.

Literature

Agterberg, F.P., 1974, Geomathematics: Elsevier, Amsterdam, 596 pp.

Agterberg F.P., and Nel L.D., 1982a, Algorithms for the ranking of stratigraphic events: Computers and Geosci. (8), 1: 69-90.

Agterberg F.P., and Nel L.D., 1982b, Algorithms for the scaling of stratigraphic events: Computers and Geosci. (8), 2: 163-189.

Brauer Hudson C., and Agterberg F.P., 1982, Paired Comparison Models in Biostratigraphy: Math. Geol., (14), 2: 141-159.

Davis J.C., 1973, Statistics and data analysis in Geology: Wiley, New York, 550 pp.

Gradstein F.M., and Agterberg F.P., 1982, Models of Cenozoic Foraminiferal Stratigraphy -- Northwestern Atlantic Margin. In: J.M. Cubitt and R.E. Reyment, eds. Quantitative Stratigraphic Correlation: Wiley, New York.

Gradstein, F.M., Agterberg, F.P., Brower, J., and Schwarzacher, W., 1985, Quantitative Stratigraphy: Unesco and Reidel Publ. Co., 600 pp.

Gradstein, F.M., Williamson, M.A., Thomas, F.P., Murney, G.M., and Stam, B., 1985, Micropaleontology of the Mesozoic-Cenzoic Sedimentary Wedge, offshore Eastern Canada. Geol. Surv. of Canada Rept., 187 pp.

Hay W.W., 1972, Probabilistic Stratigraphy, Eclogae Geologicae Helveticae, (65) 2: p. 255-266.

Williamson, M.A., in press, Quantitative biozonation of the Late Jurassic and Early Cretaceous of the East Newfoundland Basin; Micropaleontology.

Sequence Position	Fossil Number	Range	Fossil Name
1	5	0- 5	Stensioina pommerana
2	4	0- 6	Globotruncana arca
3	3	0- 6	Globotruncana stuarti
4	105	0- 5	Loxostoma gemmum
5	6	4- 6	Globotruncanella havanensis
6	16	5- 8	Rugoglobigerina rugosa
7	8	5- 8	Globigerinelloides messinae
8	13	7- 9	Globotruncana linneianna
9	12	8-10	Globotruncana stuartiformis
10	11	9-12	Globotruncana fornicata
11	14	9-12	Globotruncana cretacea
12	15	11-13	Globotruncana marginata
13	26	12-15	Globotruncana angusticarinata
14	56	12-15	Gaudryina austinana
15	20	14-17	Globotruncana coronata
16	108	12-23	Gavelinella minima
17	23	15-18	Stensioina exculpta
18	55	17-21	Globotruncana carinata
19	24	17-20	Hedbergella amabilis
20	22	19-21	Sigalia deflaensis
21	54	20-22	Globotruncana concavata
22	72	21-23	Globotruncana renzi
23	109	22-25	Coarse Agglutinated Spp.
24	110	21-25	Globotruncana imbricata
25	70	24-26	Hedbergella bosquensis
26	112	25-28	Globotruncana helvetica
27	30	25-29	Praeglobotruncana stephani
28	71	26-29	Globotruncana schneegansi
29	31	28-30	Praeglobotruncana turbinata
30	27	29-32	Rotalipora cushmani
31	131	29-32	Gavelinopsis cenomanica

Table 1. Ranked optimum sequence of the sample run, discussed in Chapter V.

Position	Fossil Pairs	Fossil Distance	Cumulative Distance	Sum Diff Z Values	Weight	S.D.
1	105- 3	.0136	.0136	.1204	8.9	.1834
2	3- 5	.0091	.0227	.1280	14.1	.1026
3	5- 4	.0584	.0810	1.1663	20.0	.0584
4	4- 6	.1107	.1917	2.3599	21.3	.0499
5	6- 8	.1651	.3569	2.2470	13.6	.1404
6	8- 16	.2048	.5616	2.8966	14.1	.1706
7	16- 13	.2432	.8049	6.2135	25.5	.0965
8	13- 12	.1327	.9375	3.3698	25.4	.0730
9	12- 14	.2873	1.2248	5.1136	17.8	.1499
10	14- 11	.0549	1.2797	1.4711	26.8	.0955
11	11- 15	.5054	1.7851	16.1465	31.9	.0847
12	15- 56	.2196	2.0047	2.2289	10.2	.1324
13	56- 26	.0481	2.0528	.2718	5.7	.2715
14	26- 20	.3530	2.4058	4.1062	11.6	.1109
15	20-108	.3876	2.7934	2.1858	5.6	.1832
16	108- 23	.9720	3.7654	.4977	0.5	999.9999
17	23- 55	.0011	3.7665	.0052	4.7	.4139
18	55- 22	.5935	4.3600	1.8888	3.2	.3458
19	22-110	.8354	5.1954	2.0430	2.4	.4111
20	110- 24	.1609	5.3564	.9680	6.0	.2589
21	24-112	.2171	5.5735	1.4250	6.6	.2770
22	112- 70	-.3006	5.2728	-2.6051	8.7	.2938
23	70-109	.0291	5.3019	.3212	11.0	.2143
24	109- 54	.0025	5.3045	.0333	13.1	.2152
25	54- 72	.1080	5.4125	1.5879	14.7	.1376
26	72- 71	.3487	5.7612	4.0611	11.6	.1745
27	71- 30	.0131	5.7743	.2116	16.1	.1270
28	30- 31	.2759	6.0503	5.0442	18.3	.0788
29	31- 27	.8785	6.9288	4.7836	5.4	.1650
30	27-131	.1588	7.0876	.7519	4.7	.1609

Table 2. Preliminary solution of scaling as given in the "distance analysis, weighted differences" - Table of the Sample Run.

New Sequence	Distance From 1st Position	Fossil Pairs	Inter Fossil Distance
105	0.0000	105- 3	.0136
3	0.0136	3- 5	.0091
5	0.0227	5- 4	.0584
4	0.0810	4- 6	.1107
6	0.1917	6- 8	.1651
8	0.3569	8- 16	.2048
16	0.5616	16- 13	.2432
13	0.8049	13- 12	.1327
12	0.9375	12- 14	.2873
14	1.2248	14- 11	.0549
11	1.2797	11- 15	.5054
15	1.7851	15- 56	.2196
56	2.0047	56- 26	.0481
26	2.0528	26- 20	.3530
20	2.4058	20-108	.3876
108	2.7934	108- 23	.9720
23	3.7654	23- 55	.0011
55	3.7665	55- 22	.5935
22	4.3600	22-110	.8354
110	5.1954	110- 70	.0774
70	5.2728	70-109	.0291
109	5.3019	109- 54	.0025
54	5.3045	54- 24	.0519
24	5.3564	24- 72	.0561
72	5.4125	72-112	.1610
112	5.5735	112- 71	.1877
71	5.7612	71- 30	.0131
30	5.7743	30- 31	.2759
31	6.0503	31- 27	.8785
27	6.9288	27-131	.1588
131	7.0876		

Table 3. The new sequence as given in the "events sorted" table of the Sample Run.
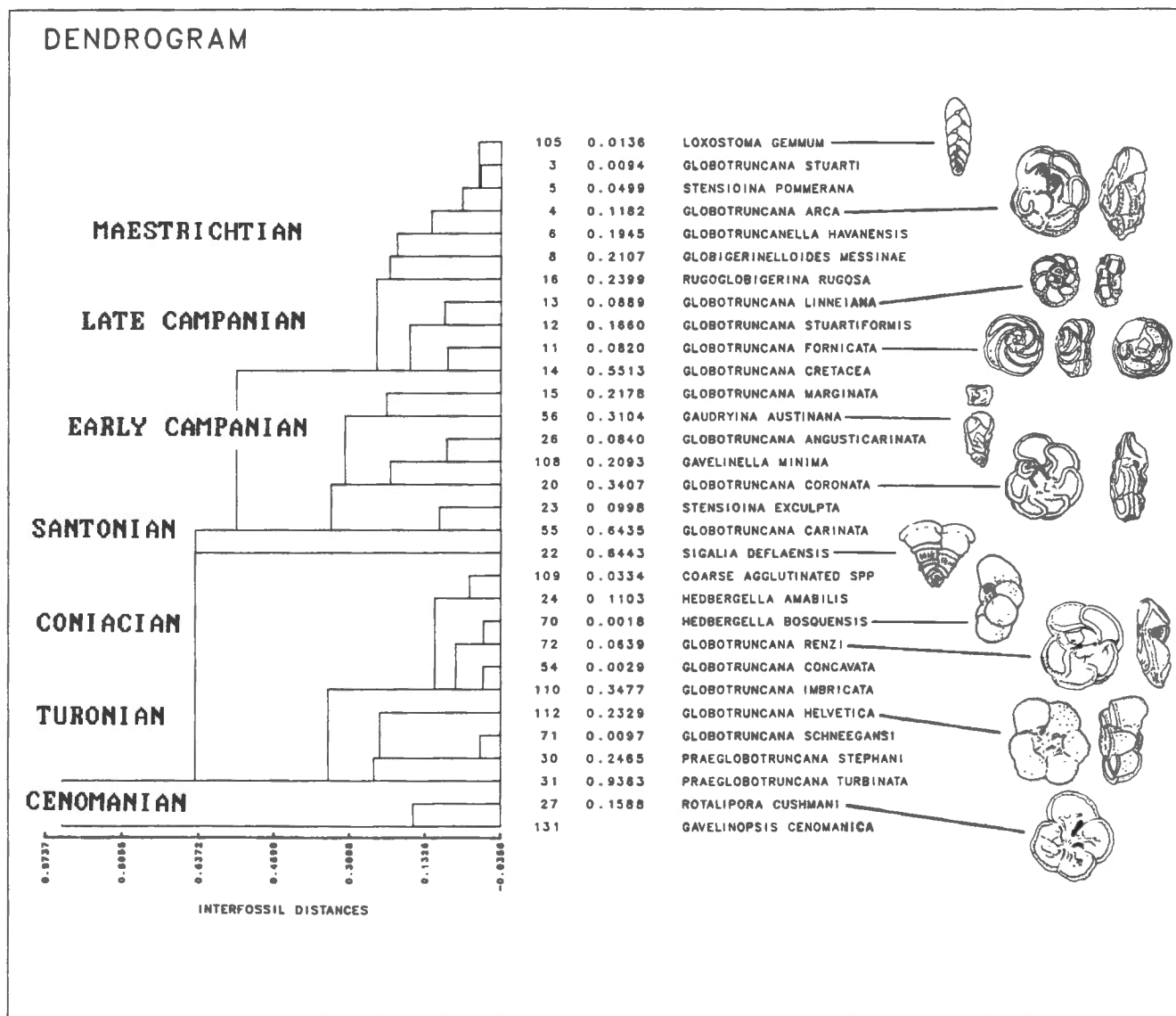
Table 4. Dendrogram of weighted interfossil distance estimates.

OCCURRENCE TABLE

NAME	NUMBER	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
LOXOSTOMA GE7MU7	105						X				X		X			X		X			
GLOBOTRUNCANA STUARTI	3			X	X	X	X				X		X								
STENSIOINA POMMERANA	5		X	X			X				X	X	X			X		X			X
GLOBOTRUNCANA ARCA	4	X		X	X	X	X	X		X	X	X	X	X				X			X
GLOBOTRUNCANELLA HAVANENSIS	6	X	X	X	X					X		X	X			X					
GLOBIGERINELLOIDES MESSINAE	8			X			X			X		X	X		X		X				
RUGOGLOBIGERINA RUGOSA	16	X	X	X		X	X				X	X				X		X			
GLOBOTRUNCANA LINNEIANA	13	X	X	X	X	X	X		X	X	X	X	X	X			X	X	X	X	X
GLOBOTRUNCANA STUARTIFORMIS	12	X		X	X	X	X				X	X	X								X
GLOBOTRUNCANA FORNICATA	11	X	X	X			X			X	X	X			X		X	X	X	X	X
GLOBOTRUNCANA CRETACEA	14			X					X	X	X	X	X	X	X	X		X	X	X	X
GLOBOTRUNCANA MARGINATA	15		X					X	X	X	X	X	X	X			X	X		X	X
GAUDRYINA AUSTINANA	56	X							X								X	X			X
GLOBOTRUNCANA ANGUSTICARINATA	26			X							X		X		X		X				X
GAVELINELLA MINIMA	108							X	X	X									X	X	
GLOBOTRUNCANA CORONATA	20		X	X					X		X		X		X		X				X
STENSIOINA EXCULPTA	23	X		X								X	X			X					
GLOBOTRUNCANA CARINATA	55	X							X		X	X				X					X
SIGALIA DEFLAENSIS	22			X						X		X					X				X
COARSE AGGLUTINATED SPP	109							X	X	X	X									X	X
HEDBERGELLA AMABILIS	24		X	X								X				X			X	X	X
HEDBERGELLA BOSQUENSIS	70		X							X				X		X	X			X	X
GLOBOTRUNCANA RENZI	72		X					X	X	X	X	X		X	X	X					X
GLOBOTRUNCANA CONCAVATA	54	X	X						X		X	X	X	X			X	X	X		X
GLOBOTRUNCANA IMBRICATA	110								X	X							X		X	X	
GLOBOTRUNCANA HELVETICA	112								X	X	X						X		X	X	
GLOBOTRUNCANA SCHNEEGANSI	71		X						X	X	X			X						X	X
PRAEGLOBOTRUNCANA STEPHANI	30			X	X				X	X	X		X				X		X	X	X
PRAEGLOBOTRUNCANA TURBINATA	31			X					X	X	X					X		X		X	X
ROTALIPORA CUSHMANI	27			X									X		X	X			X	X	
GAVELINOPSIS CENOMANICA	131								X			X		X	X				X	X	X

Table 5. Occurrence Table.

PETREL A-62

```
         I    4   14   15   13   72   27  131
---------I----------------------------------
     5 I
     4 I XX
     3 I
   105 I
     6 I
     8 I
    16 I
    13 I                 XX
    12 I
    11 I
    14 I      XX
    15 I           X X
    26 I
   108 I
    56 I
    20 I
    23 I
    55 I
   109 I
    24 I
    22 I
    54 I
    72 I                      XX
   110 I
    70 I
  ·112 I
    30 I
    71 I
    31 I
    27 I                           XX
   131 I                                XX
```

Table 6.   Scattergram or bivariate plot for Petrel A-62.
           Scaled optimum sequence on the vertical axis.

Petrel A-62		Cum. Dist.	2nd Order Difference
Globotruncana arca	4	0.0810	
Globotruncana cretacea	14	1.2248	−0.5835
Globotruncana marginata	15	1.7851	−1.5406
Globotruncana linneiana	13	0.8049	5.5879**
Globotruncana renzi	72	5.4125	−3.0913*
Rotalipora cushmani	27	6.9288	−1.3575
Gavelinopsis cenomanica	131	7.0876	

* – Greater than 95% Prob. Event Out of Position
** – Greater than 99% Prob. Event Out of Position

Table 7. Normality test output for the Petrel A-62 well of the Sample Run (see Chapter V).

NORMALITY TEST

COMPARISON OF OBSERVED AND EXPECTED OCCURRENCES OF SECOND ORDER DIFFERENCE VALUES

CLASS NO. OBSERVED   EXPECTED DIFFERENCE   DELTA

CLASS NO.	OBSERVED	EXPECTED	DIFFERENCE	DELTA
1	20	21.100	-1.100	.037
2	21	21.100	-.100	.000
3	30	21.100	8.900	2.402
4	13	21.100	-8.100	1.989
5	23	21.100	1.900	.109
6	17	21.100	-4.100	.510
7	29	21.100	7.900	1.892
8	18	21.100	-3.100	.291
9	18	21.100	-3.100	.291
10	22	21.100	.900	.025

CHI-SQUARED =   7.547

* -GREATER THAN 95% PROBABILITY THAT DIFFERENCE IS NOT ZERO
** -GREATER THAN 99% PROBABILITY THAT DIFFERENCE IS NOT ZERO

Table 8.   Normality Test - Approximated Chi Square Test.

Petrel A-62		Dictionary Number
Sample	Event	
2970'	Globotruncana arca	4
3090'	G_____ cretacea	14
3360'	G_____ marginata	15
3390'	G_____ linneiana gr.	13
4150'	G_____ aff. renzi	72
4320'	Rotalipora cushmani	27
4350'	R_____ appenninica	33
	Praeglobotruncana sp.	170
4408'	Gavelinopsis cenomenica	131
4466'	Favusella washitensis	169

Table 9.   Format of raw data listing of the Petrel A-62
           well, used in the Sample Run. The events listed
           are highest stratigraphic occurrences.

Tabulation of Event Occurrences;																			
No. of Wells	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
No. of Events	21	9	7	8	6	6	5	2	3	3	1	1	1	2	0	0	1	0	0	0
Cum. of Events	76	55	46	39	31	25	19	14	12	9	6	5	4	3	1	1	1	0	0	0

Table 10.   Second preprocesing table of the Sample Run; this table tabulates the event occurrences and is used to determine k_c values.

1	Globotruncana contusa	54	Globotruncana concabata
2	Fructicose Pseudotextulariaptus	55	Globotruncana carinata
3	Globotruncana stuarti	56	Gaudryina austinana
4	Globotruncana arca	57	Globotruncana ventricosa·
5	Stensioina pommerana	58	Globigerinelloides aspera
6	Globotruncanella havanensis	59	Rugoglobigerina rotundidorsata
7	Pseudotextularia elegans	70	Hedbergella bosquensis
8	Globigerinelloides messinae	71	Globotruncana schneegansi
9	Globorotalites michelinianus	72	Globotruncana renzi
10	Arenobulimina americana	95	Globotruncana falsostuarti
11	Globotruncana fornicata	104	Globotruneana rosetta
12	Globotruncana stuartiformis	105	Loxostoma gemmun
13	globotruncana linnaiana	106	Arenobulimina d'orbignyi
14	Globotruncana cretacea	107	Kyphopyxa christneri
15	globotruncana marginata	108	Gavelinella minima
16	Rugoglobigerina rugosa	109	Coarse agglutinated spp.
17	Globorotalites aff. multiseptus	110	Globotruncana imbricata
18	Praebulimina sp.	111	Globotruncana primitiva
19	Aragonia materna kugleri	112	Globotruncana helvetica
20	Globotruncana coronata	113	Globotruncana sigali
21	Reussella szajnochae	114	Globotruncana marianosi
22	Sigalia deflaensis	115	Lingulogavelinella turonica
23	Stensoinia exculpta	116	Hedbergella paradubia
24	Hedbergella amabilis	128	Gavelinella tourainensis
25	Hedbergella spp.	129	Praeglobotruncana hagni
26	Globotruncana angusticoronata	130	Praeglobotruncana difformis
27	Rotalipora cushmani	139	Gaudryina pyramidata
28	Rotalipora greenhornensis	141	Hedbergella simplex
29	Rotalipora deckeri	140	Marsonella oxicona
30	Praeglobotruncana stephani	168	Globotruncana gansseri
31	Praeglobotruncana turbinata	169	Favusella washitensis
32	Grandes hedbergelles	170	Praeglobotruncana sp.
33	Rotalipora appenninica	173	Planulina taylorensis
34	Praeglobotruncana delrioensis	174	Epistomina stelligera alveolata
35	Dorothia aff. filiformis	175	Vaginulina texana
48	Hedbergella planispira	178	Neoflabellina rugosa
50	Agglutinated spp abundant	184	Petrel limestone*
53	Globotruncana conica		

Table 11. Dictionary of the 76 Late Cretaceous events of the Sample
Run. In the computer output the complete list of 184
Jurassic-Cretaceous events will be printed out.

ORIGINAL SEQUENCE DATA

BONAVISTA-CGG
   4   -12 -13 -53 -11 -6 -16   54 -55 -17 -56 -23

CIMBERLAND B55
  21   -16   -2 -13 -11 -6   -5   70 -24 -71 -20 -54 -72

ADOLPHUS O-50
   1   -2   -3   -4   -5 -6   -7   -8 -9 -10 -11 -12 -13 -14 -15 -16 -17   18   19 -20
 -21   22 -23 -24 -25 -26 184   27 -28 -29 -30 -31 -32   33

BLUE H-28
   4   -13 -12 -57   -3 -21   -7   -58 -6 -59   33 -30 -48 -50 -35

A GABRIEL C-60
   4   -11 -12 -13 -57 -95   -3   16

HARE BAY E-21

   6   -16 -53   -4   -3 -12 -13 -104 -7   -5 -105   -8 106 107

EGRET K-36
   4   55 108   72 -30 112 -111   71 -15 -109 -128 -110 -113 -114   31 -129 -130 -115 -54

HIBERNIA O-35
 139   -10 -16   -6 -14 -13   -4   20 -11 -55 -15 -56 -140 112   30 -71 -113 -141 131 -31
-108 -109

EGRET N-46
   9   -8 -13   -4   -6 11   14   22 -15 108 -54 109 -110 -72 -111 112 -71 -113 -25 -114
 -70   -31 -115 -116   30

PUFFIN B-90
   3   -4 -105   -5   12 57 -11   13 14   20 -15 -26   23   54 -55   72 -113 109

KILTIWAKE P-11
   4   -6 -13   -5   -1 -16   8   168 12 -14   11 -15   23 -24   54 -22   72 -30 131 -34
  27   -28 169

HERMINE E-94
   4   -6   -3 -168 -105 -5   -7   16 -8   13 -14   26 -15   12   20   70   54   71 170

PETREL A-62
   4   14   15   13   72 27   33 -170 131 169

HERON H-73
   8   14 -11 -107   17 -26   23   -20 24   70   55   72 184   27 -31 131 169

BITTERN M-62
  16   105   -5   6   14 22   54   72

KIDER M-75
   8   14 -107 -173 174 15 -175   56 -13   9   20 -26 -11 -70   54 112 -110 -30 -31

JAEGER A-49
   4 -105 -16   -5 107 -173 -15   -11 -13 -178 -174 175   56   54

CORMORANT K-83
  14   -13 -11 112 -30 -31 -110 -115 -24 -108   27 -28 -131

MURRE G-67
  11   -13 -15 -14 108 -109 -110 115 -70   30 -31 112 -71 -24 131   27 -28

OSPREY H-84
 173   -10   -9 -17 -11 -13   -4   -12 -5   26 -57 -53 -14   55 -56 -109   20 -15   72 -71
  -54   -22 -70 -24 111 -30 131

Table 12. List of coded sequences for the Sample Run.