

# **Neogene biochronology of Antarctic diatoms: A comparison between two quantitative approaches, CONOP and UAgraph**

by

Federico Galster <sup>1</sup>, Jean Guex <sup>1</sup> and Oyvind Hammer <sup>2</sup>

<sup>1</sup> Department of Geology, Anthropole, CH-1015 Lausanne, Switzerland

<sup>2</sup> Natural History Museum, University of Oslo, Norway

## **Abstract**

A recent quantitative biochronological study by Cody et al. (2008) integrates comprehensive diatom biostratigraphy, magnetostratigraphy, and tephrostratigraphy from 32 Neogene sections around the Southern Ocean and Antarctic continental margin. A recent method, known as Constrained Optimization (CONOP), which can be viewed as a multidimensional version of graphic correlation, is applied to that very interesting database.

The goal of the present paper is to discuss some theoretical aspects of quantitative biochronology and to compare the constrained optimization with the deterministic method called Unitary Associations (UAM), a graph theoretical model. We illustrate the fact that the UAM is an extremely powerful and unique theory allowing an in depth analysis of the internal conflicting inter-taxons stratigraphic relationships, inherent to any complex biostratigraphical database.

# 1 Introduction

Cody et al. (2008) recently produced a convincing demonstration that quantitative stratigraphic approaches can produce results with a much higher resolution potential than empirical zonations. Their approach was based on the method known as “Constrained Optimization”, using the program CONOP (Sadler 2006), and provided a splendid example to be compared with other methods.

The goal of the present paper is to compare the constrained optimization with an alternative quantitative tool, the Unitary Associations, a graph theoretical model, and explain the differences between the outputs. It is a common belief among non specialists that all quantitative tools in stratigraphy are more or less equivalent but we do not agree with that idea. The present paper should help the user of quantitative biostratigraphical tools in their choice of one or the other method.

A short description of the two approaches is given below.

## 2 Conop and UAs: a brief description

### ***2.1 The UA method.***

The Unitary Associations (UAs) method is designed for the construction of concurrent range zones using a fully deterministic approach. The basic idea of the method is to construct a discrete sequence of coexistence intervals of species. Each interval, corresponding to one UA, is of minimal duration while consisting of a maximal set of intersecting ranges. Each UA is characterised by a set of species allowing its identification in the stratigraphic sections.

The basic steps of the method are as follows. The data are compiled into a presence-absence matrix, with samples in rows and taxa in columns. From these data, maximal sets of mutually

co-occurring species (maximal cliques) are constructed. Stratigraphical superpositions of maximal cliques are then inferred from the observed superpositional relationships between the taxa they contain. The longest possible sequence of superposed maximal cliques is then used to construct a sequence of UAs. Finally, the original samples are assigned to UAs whenever possible and are thus stratigraphically correlated.

The difficult part of the UA theoretical model consists in finding and resolving what we call conflicting stratigraphic relationships, i.e. cyclic structures in the graphs representing the inter-taxa stratigraphic relationships on one side and the strongly connected components in the graph representing the sequential relationships between the maximal cliques ( $G_k$ ) on the other side. A full review of the method of Unitary Associations is out of the scope of the present paper and can be found in Guex (1991).

The program used in the present paper is the latest version of UA-graph (Hammer, Guex and Savary 2009, available at <http://folk.uio.no/ohammer/uagraph>) which is an improved version of the Biograph program (Savary and Guex 1991).

## ***2.2 The Constrained Optimization method***

The Constrained Optimization method used by Cody et al. (2008) was originally designed to produce automated graphic correlation in a multidimensional space (Kemple et al. 1995) and sequences of events by means of the program named CONOP. The first rule used by the authors is to satisfy the observed interspecies coexistences, as in the UA method. One particular type of contradictory situation treated by CONOP is given in Fig. 1.

FIG 1

The second criterion is that the computed solution must satisfy the condition that adjusted FO and LO horizons lie at or below and at or above the observed FO and LO respectively. The latest version of CONOP program (Sadler 2006) offers multiple choices and adjustments of a variety of different parameters, such as specifying the measure of misfit and comparing the penalty of several different sequences

CONOP considers the problem of stratigraphic correlation as NP-complete (Dell et al. 1992; see remark below). Sadler (2006) notes that the CONOP program minimizes the simplifying assumptions and maximizes the flexibility of the choice of measures of fit between solutions and the data. This is achieved by inverting the solution process. Instead of building a solution from the data (as in the UA method), CONOP works through a series of iteratively improved guesses about the solution. Each guess is compared with the data; the misfit between the solution and data guides the next guess, a process called "inversion" by geophysicists. Cody et al. (loc.cit.p.xxx ) minimize the corrections of local ranges by using an option called LEVEL as penalty function, i.e. the minimal value upon which the simulated annealing is calculated (details in Sadler 2006). Our present discussion is limited to this particular choice.

When judiciously configured by an experienced user, the program may be tailored to contract or extend local ranges or limit the use of the coexistence constraint by eliminating some rare co-occurrences (i.e. observed in only one sample) during the computation. Again, a full description of the model is outside the scope of the present paper and the reader is referred to Kemple et al. (1995) and Sadler (2006) for details.

In summary, the main differences between UAs and CONOP is that the goal of the first method is to construct concurrent range zones and the second construct sequences of datums, resulting into Interval Zones. Note that UA constructions are based on samples contents and their mutual stratigraphic relationships (i.e. the biostratigraphic graph  $G^*$ ). Sequences of

datums established by means of UAs consist in non diachronous datums only which are constructed in a second run of the program, after a semi-empirical selection. During the correlation process, such datums are considered as if they were interval of uncertainty, like the total ranges of the taxa calculated by the program.

CONOP on its side is exclusively oriented towards a seriation of the events based on inter-stratigraphic sections comparisons (i.e. graphic correlation philosophy).

We can also notice that the two methods will reach similar conclusions when applied to very complete sets of empirical information relative to inter-taxa coexistences.

### **3 Technical remark about the NP-completeness of the problem**

In our theoretical model, the stratigraphic correlation problem is not addressed as an NP-complete problem but as a problem of seriation of the observed samples. Sequences of first and last occurrences (FOs and LOs) are only treated when these events are not diachronous in the different localities. This explains the important differences existing between the computing time used by UAglyph and that used by CONOP. In our program, the number of computing operations is linearly proportional to the number of samples and to that of the taxa involved. In CONOP the growth is exponential but can be limited at the expense of precision.

### **4 The database**

One major interest of the Cody et al. (2008) paper is that the complex problem which was originally addressed, the quantitative study of 116 diatom species biostratigraphy, magnetostratigraphy and tephrostratigraphy from 32 Neogene sections around the Southern Ocean and Antarctic continental margin has been analysed by the best specialists on the

CONOP program, producing a result which can be interpreted as one of the best possible result obtainable by means of Constrained Optimization.

Another major interest of Cody's database is that the original taxonomy was carefully revised and homogenised and that the doubtful local range extremities (reworking and downworking, contaminations) were consciously eliminated.

#### **4.1 UA processing of the data**

All the data considered here are strictly taken from Cody's database (Online Appendix 2 in Cody et al. 2008) which is considered, for theoretical reasons, as exact.

The goal of our application of the new UA program UAGraph (Hammer, Guex and Savary in PAST) is to produce an *integrated* solution given the complex input including the inter-species coexistences, selected moderately diachronous first and last occurrences and paleomagnetic data. The database used for the computation, as noted above, is strictly derived from Cody's compilation and given in Online Appendix 1 (1-A: local ranges; 1-B: taxa dictionary; 1-C: sections codes).

The paleomagnetic data (base/resp. top of the recognized reversals) have been included into the computed database to force an order on some contradictive (cyclic) relationships. As the truly oldest (resp. youngest) part of the considered reversals are not necessarily recorded in the local stratigraphic sequences, these events often display an apparent range which can span several distinct UAs.

Concerning the "datums", we should recall that stratigraphers usually say that a given datum, or first (resp. last) occurrence, is diachronous when its chronological position varies from one stratigraphic section to the other. Knowing how many zones a datum jumps from one place to the other allows us to quantify its diachronism, i.e. if a FO jumps one zone we say that its diachronism is equal to 1, if it jumps two zones, its diachronism is equal to 2 etc. Generally

speaking, FOs and LOs sequences can only be used when these events are proven to be non diachronous. Moreover, the only ones which are useful for correlating stratigraphic sections are those widely distributed over large geographic areas. In other words, a sequence of FOs or LOs observed in only one section is useless for correlation. In our treatment of Cody's data, we have introduced about 30 datums (FOs) showing a low apparent diachronism (list in Fig. 2). Again we note that the true evolutionary appearance (also called First Appearance Datum or FAD) of a taxon is very rarely recorded in any stratigraphic sequences (the probability to record such an event is equal to zero) and even the "best" first occurrences show an apparent range that can span several distinct UAs, due to regional diachronism.

#### **4.2 Remark about reworking**

In some cases, the diachronism of datums can be due to reworking or downworking. To detect such situations, we can proceed as follows. For each UA recognized in any given section, we compile the list of the taxa present in all beds assigned to the UA under consideration. This allows us to recognize the physical absence (but virtual presence) of certain taxa in some beds assigned to specific UAs (see for example the discontinuous taxa in Fig. 2).

In the final computed UA range chart, UAgraph denotes the chronological absences of taxa from some UAs by an x (=discontinuity) in the column corresponding to the total range of the considered taxon.

For example, in our 92 UA range chart given in Fig. 2 (see also Fig 5a), we note that the highly discontinuous taxon *Fragilariopsis pusilla* (66) could be reworked in UA 52, and that taxons *Thalassiosira eccentrica* (107) and *Thalassiosira jacksonii* (113) could be downworked in UAs 46-47. To be certain of a case of reworking, we should re-examine the original samples but this is of course beyond the scope of this paper.

The usual input of the data into UAGraph is given in the format "datum", i.e. the local range of each taxon is given as a triplet of codes: taxon, level of first occurrence and level of last occurrence. Such input obviously hides local discontinuities in the stratigraphical distribution of the taxa. However, when the input of the data into UAGraph is given as lists of taxa present in the different paleontological samples (= "sample" format), the significant local discontinuities are also illustrated in the final range chart and reinforce the efficiency of the above procedure.

FIGURE 2:

## 5 Standard use of a range chart for correlation

\*\*\*\*

Most micropaleontologists use "Interval Zones", i.e. biochronologic units where zones are defined by first or last occurrences (FOs, resp. LOs) of taxa. In other words they apply zonal range charts by using the sequences of events (FOs and LOs) represented in these range charts. When the basic sequences of datums are based on synchronous first and last occurrences (i.e. the local sequences are identical to the sequence established in the Interval Zone range chart), the process of correlation is straightforward and poses no problems. However, if the local events (FOs and LOs) are diachronous, which is very common, the standard sequences of datums cannot be applied to produce precise correlations.

An alternative way to use a standard range chart is to apply the following basic reasoning: If bed B contains taxon  $i$ , then B was deposited during the existence interval of taxon  $i$ . If B contains  $i$  and  $j$ , then we conclude that B was deposited during the coexistence interval of taxon  $i$  and  $j$  ( $i,j$ ). If B contains taxons  $i,j,k \dots,n$ , then it was deposited during the coexistence interval of the  $n$  taxa. If  $n$  is maximum, then the coexistence interval has a minimal duration. Note that the absolute duration (i.e. chronometric) of the coexistence interval is obviously

unknown (detailed discussion in Guex 1991; see also Escarguel and Bucher 2004). A discrete sequence of distinct and ordered coexistence intervals of minimal duration is called a discrete biochronologic time scale, the units of which are concurrent range zones. Unitary Associations produces such discrete scales. Note that this reasoning can be applied to *any* range chart, provided that the range chart contains all the data that have been observed in the stratigraphic sections. If that condition is not satisfied, the range chart is incomplete and needs to incorporate new information or it is merely false. Note that the correlation reasoning explained above does not apply to the Cody et al. correlation chart (their Fig. 4b) where most of the lines of correlations generated by the "events" relationships are converging to one point, suggesting reduced sedimentation in some of the sections. This is due to the fact that the initial observations are parsimoniously adjusted according to the expected sequence of events calculated by CONOP (Cody et al. loc. cit. p. 109). Such convergences do not appear in the UAgraph correlation output.

## **6 Cody's and UA outputs: Criteria of comparison between the UAgraph and CONOP**

The main constraint in the UA is to preserve a maximum of observed and reproducible superpositions while honouring the observed co-occurrences in the output. That condition was demonstrated long ago to be the most important constraint in biostratigraphic analysis (Guex 1977, 1979). To honour that condition, the UA algorithm is oriented towards the systematic destruction of cyclic structures present in the graph representing the inter-taxa stratigraphic relationships. The basic conditions adopted in the CONOP program are similar but the approach for reaching an optimal solution is totally different from the one adopted for UAs.

For reasons explained below, the details of the FOs (resp. LOs) relationships represented in Cody's original range chart (loc.cit. Fig. 6) are not reproduced here. Fig. 4a is a presentation of Cody's range chart (considered as a single section) run through UAGraph and reduced to its maximal intersections (see Fig.3). That reduced solution consists in 37 maximal intersections (Fig.4a) which can be compared with the UAGraph output using all Cody's sections as input, producing a sequence of 41 UAs (Fig.4b).

FIG 3

Fig 4

The correlation table (= "fence diagram" in CONOP terminology) resulting from the application of Cody's range chart following the above reasoning on concurrent range zones is given in the Online Appendix 2. The reproducibility table of Cody et al. maximal intersections (Fig 5-c) shows the stratigraphic sections (columns 1-32) and the different units (corresponding to rows 1-37 in Fig.4a) resulting from that correlation and constructed as explained above (Fig.3). It is correlated with its equivalent where the diatom original database is calculated by means of the UAs (Figs. 4b and 5b). That 41 UAs solution is itself correlated with our 92 UAs solution (Figs. 2 and 5a), based on the original diatom local ranges together with some FOs and LOs selected for their reliability, including paleomagnetic and ash beds data.

The two tables of Figs. 4b and 4c show that the two approaches produce similar results. We note however that, in this case, UAGraph needs to construct 456 virtual edges only instead of 550 by CONOP (i.e. about 20% less) to transform the initial biostratigraphic graph into an

interval graph (see Guex 1991, Chapter 4 for technical explanation). That difference can be explained as follows.

As noted above, the most difficult part of the UA theory concerns the treatment of conflicting stratigraphic relationships, i.e. cyclic structures in the graphs representing the inter-taxa stratigraphic relationships on one side and the strongly connected components in the graph (noted as  $G_k$  = graph of cliques in UAgraph) representing the sequential relationships between the maximal cliques on the other side.

Fig 5

One simple example of such conflicting data is illustrated in Fig 6, showing two samples, 1 and 2, which have a conflicting stratigraphic relationship because two inter-taxa superpositions relating the samples are oriented in an inverted order: A is simultaneously above and below B. Such relationships must be eliminated by destroying one of the conflicting superpositions, i.e. by replacing it by a virtual coexistence (=virtual edge in the graph  $G^*$ ). One of the basic goals of the method is to minimize the number of superpositions which are replaced by what is called “virtual coexistences” (i.e. inter-taxa coexistences which are real in time but are not observed physically in the paleontological samples). Doing so, we optimize the preservation of their reproducibility.

FIG 6

Note that one of the outputs of the program UAgraph gives an enumeration of all the cycles

of length 4 (noted as Z4) present in the biostratigraphic graph under study. In some databases like the one compiled in the large Memoir edited by Peter Baumgartner (1995) on the stratigraphy of the Jurassic and Cretaceous radiolarians, there are almost one million such cycles.

The situations treated by CONOP are different, because this theoretical model considers that two taxa  $i$  and  $j$  display a conflicting distribution if they are observed in sequence in one locality and are co-occurring in another (see Fig 1). Such apparent contradictions are not necessarily belonging to cyclic relationships, such as described in Fig. 6. When such a situation is encountered, CONOP is designed to extend the range of  $i$  upward or  $j$  downward. Here we can note that the best-fit interval values used in CONOP to characterize the uncertainty of events assignments, are generally unrelated to the amount of implication of the taxa within cyclic relationships, such as Z4 and discontinuous stratigraphic distribution.

In contrast, our optimization procedure is designed to preserve a maximum of observed superpositions (= reproducibility of the arcs in  $G^*$ ) (output "Contradictions" in UAGraph) and it operates as follows:

For each pair  $ki, kj$  of maximal cliques of  $G^*$  showing a contradictory stratigraphic relationship, we define two sets of arcs, A and B, where A is the set of arcs that links the elements of  $ki$  to those of  $kj$  (in the direction  $ki \rightarrow kj$ ) and where B is the set of arcs (of the opposite orientation) that links the elements of  $kj$  to those of  $ki$ . To each set A and B we attribute a value  $V(A)$  (resp.  $V(B)$ ) equal to the sum of the individual reproducibilities of the arcs belonging to A (resp. B) added to the number of arcs of A (resp. B).

If  $V(A) > V(B)$  we say that the clique  $kj$  is "located above"  $ki$  (keeping in mind that this is an abuse of language and that the superposition is conflicting).

If  $V(A) = V(B)$  we say that the stratigraphic relationship between  $ki$  and  $kj$  is undetermined.

This procedure is equivalent to making a global (and no longer individual) search for the forbidden configurations of  $G^*$ .

An example of application to the diatom database is given in Fig. 7, showing the conflicting relationships between samples 5.35 (section 5, bed 35 = sample  $i$ ) and 28.18 (= sample  $j$ ).

Fig.7

In this example, we see that there are 4 arcs connecting  $i$  (sample 5.35) to  $j$  (sample 28.18) with a total reproducibility (i.e. the number of times the superposition was observed in the sections) of 15 (total = 19) and 3 arcs connecting  $j$  to  $i$  with a total reproducibility of 18 (sum = 21). The order between  $i$  and  $j$  calculated in the initial  $G_k$  graph is  $i \rightarrow j$ . Note again that the case where the values are identical is considered as undetermined.

Coming back to the above discussion, we note that the 92 unit UA range chart given in Fig. 2 is difficult to compare directly with the reduced CONOP's range chart (Fig.4a) because the UA solution is calculated from much more information, including maximal intersections of species ranges, paleomagnetic data and a collection of datums (all taken from Cody's data base). However, a careful comparison with CONOP's original chart (loc cit Fig. 6) shows that two intervals seem to have a greater resolution in CONOP's output.

1) The interval between UA 68 to UA 84, containing the paleomagnetic interval C2An-1ntop (2.58 Ma) to C2An.3nbase (3.6 Ma) has no precise resolution in the UAgraph output. We note in particular that paleomagnetic events C2An.3n top to C2An.1n top overlap

without being distinguished from each other. The reasons of this are discussed below and the same remark applies to paleomagnetic events C5An.1n base and top.

During the interval 2.58 Ma to 3.6 Ma, CONOP's composite sequence versus composite section curve (loc. cit. Fig. 5) presents a sequence of 29 increments characterized by FOs/LOs relationships and paleomagnetic events. That sequence is partly based on an interpolation between two paleomagnetic tie-points respectively located at 2.58 Ma and 3.6 Ma. To explain that problematic situation we have constructed and analyzed the graph of the relationships between these events with the result illustrated in Figs. 8a and 8b.

Fig. 8a shows what we call the trivial virtual edges between the events, generated by specific sequences which are inverted from place to place: such coexistences are generated by the stratigraphic relationships where an event  $i$  is above an event  $j$  in one place and when the reverse relationship is observed in another section. Note that in this case, the virtual edges of the graph correspond to strictly real, but physically not observed, coexistences. A consequence of this is that computing maximal cliques is perfectly allowed and provides a true sequence of non contradictory arcs. Fig 8b shows the oriented graph of the 29 events with the cycles  $C_n$ , including paleomagnetic data, and Fig. 9 represents a selection of FOs local stratigraphic sequences in the different sections (1 to 32) and the contradictory age (Ma) assigned by CONOP, explaining the origin of the cycles  $C_n$  in the graph of Fig.8b.

2) The first 23 FOs older than 18 Ma, located at the base of Cody's range chart:

When analysing the Cody et al. problem, we realized that we were unable to reconstruct the initial sequence of 23 FOs (listed in online Appendix XXXX) established by CONOP (Cody et al. table 3) where those FOs appear to be regularly separated by increments of 0.2 Ma. The difference between our sequence and the CONOP one is due to the fact that cycles  $C_n$  and trivial virtual coexistences, illustrated in Fig. 10, are not taken into account by this program, even if they should be treated before the final output.

This implies the impossibility for UAGraph to compute the "true order" of these events because the order in which such cyclic events appear in any sort of calculated output is not reliable.

Fig.8

Fig 9

Fig 10

## 7 Average Range Model

CONOP can be set up to relax the constraint against local range contraction and force the coexistences to be displayed in the output. As well noted by Cody et al. (loc cit) this is likely to contract many legitimate in situ range-ends arbitrarily and significantly underestimate the true temporal range of most taxa. According to Cody et al., such contractions provide a minimum to complement the Total Range Model's maximum estimate.

Comparing Cody's Average Range Chart with their original database by means of the UAGraph tools ("Biostratigraphic graph" and "Compare G\*") we note that 670 coexistences were destroyed and 2 arcs have been reversed (See Online Appendix 3-B). Such results can be compared with the ones obtained by the probabilistic approaches discussed by Guex (1991) and Boulard (1993).

## 8 CONCLUSIONS

The main difference between the UA theoretical model and CONOP resides in the fact the latter is essentially based on an algorithm (Simulated Annealing , see Sadler 2006 for details) which doesn't allow to treat the inter-FOs (resp. LOs) cyclic relationships.

In other words, the inter-events cycles have to be calculated separately to know which of CONOP's FOs (or LOs) partial sequences are chronologically meaningful.

As an example we can consider the cyclic relationships linking the events listed in Figs. 8 and 9. All those taxa are considered as chronologically meaningful by CONOP (see the corresponding list in Cody, loc.cit. Table 3, p.102).

In our approach such contradictions are resolved by considering the events belonging to a given cycle as equivalent.

By contrast, the Unitary Associations theoretical model and our new program UAGraph (Hammer, Guex, Savary in PAST) are designed to produce a complete analysis of the complexity and structure of the paleontological data used to make correlations, and the computation is extremely rapid (a few minutes for the largest problems treated by means of the UAs).

In particular, the following points are very important:

- 1) Analysis of the structure of the data and of its internal contradictions (list of cycles Z4 and Cn in G\* and strongly connected components in Gk).
- 2) Individual comparison of the fossiliferous horizons showing conflicting stratigraphic relationships.
- 3) Display of the chronological discontinuities of the taxa distribution in the Range Chart.
- 4) Variety of tools allowing the analysis of the diachronism of the datums and the search for reworkings.
- 5) Analysis of the purely oriented graph of FOs (resp.LOs) containing all the virtual edges generated by the cross-over  $a>b$  and  $a<b$ .
- 6) Display of all the single co-occurrences associated with multiple superpositions in other localities.

To conclude, we note that the power of resolution of the UA method is contained in the data and not in the method. This is not true for every quantitative method because some of them systematically produce total ranges which are truncated, or arbitrarily constructed and non reproducible sequences of FOs and LOs that are implicated in cyclic and conflicting stratigraphic relationships. The UA theoretical model is unique in providing a full analysis of the internal complexity of any biochronological problem. In this it differs from all other quantitative methods available today.

## References

- Agterberg, F.P., Gradstein, F.M., 1999. The RASC method for ranking scaling of biostratigraphic events. *Earth Science Reviews* 46, 1–25.
- Baumgartner, P.O., O'Dogherty L., Gorican S., rquhart E., Pillevuit, A, De Wever P. eds. 1995. Middle Jurassic to Lower Cretaceous Radiolaria of Tethys. Mém.Géologie Lausanne 23, 1-1172.
- Boulard (1993).
- Cody, R.D. , Levy, R.H., Harwood, D.M., Sadler, P.M., 2008. Thinking outside the zone: High-resolution quantitative diatom biochronology for the Antarctic Neogene. *Palaeogeography, Palaeoclimatology, Palaeoecology* 260, 92–121.
- Dell, R., Kempe, W.K., and Tovey, P., 1992. Heuristically solving the stratigraphic correlation problem, in: *Proceedings of the Institute of industrial Engineers First Industrial Engineering Research Conference*. 1, 293-297
- Edwards, L.E., 1984. Insight on why graphic correlation (Shaw's method) works. *Journal of Geology* 92, 583–587.
- Guex, J., 1977 Une nouvelle méthode d'analyse biochronologique. *Bull. Géol. Lausanne* No 224, 309-322

Guex, J., 1979 Terminologie et méthodes de la biostratigraphie moderne. Bull. Géol.

Lausanne 234, 169-216

Guex, J., 1991. Biochronological Correlations. Springer Verlag.

Hay, W.W., 1972. Probabilistic stratigraphy. *Eclogae Geologicae Helveticae* 65, 255–266.

Kemple, W.G., Sadler, P.M., Strauss, D.J., 1995. Extending graphic correlation to many dimensions: stratigraphic correlation as constrained optimization. In: Mann, K.O., Lane, H.R.,

Kirkpatrick, S., Gelatt, C.D., and Vecchi, M.P., 1983. Optimization by simulated annealing. *Science*. 220, 671-689.

Sadler, P.M., 2006. Constrained optimization approaches to the paleobiologic correlation and seriation problems: Part 1 (A Users Guide to the CONOP Program Family) and Part 2 (A Reference Manual to the CONOP Program Family).

Sadler, P.M., Cooper, R.A., 2003. Best-fit intervals and consensus sequences. In: Harries, P.J. (Ed.), *Approaches in high-resolution stratigraphic paleontology*. Kluwer Academic Publishers, Netherlands, pp. 49–94.

Shaw, A.B., 1964. *Time in Stratigraphy*. McGraw Hill, New York. 365 pp.

#### Acknowledgements

Profs Hugo Bucher and Pete Sadler made useful comments on our manuscript. This paper was published in the frame of Swiss NSF project 20021-124375-1.

## CAPTIONS

Fig. 1 A typical conflicting situation considered and treated by CONOP (after Sadler and Cooper 2003). Such cases are not considered as conflicting in the UA method.

Fig. 2 UA range chart calculated by UAgraph, including paleomagnetic data and selected FOs and LOs (data from Cody et al. 2008). Crosses (x) mark discontinuities (see text).

Arrows indicate paleomagnetic data. The regional reproducibility of the 92 units is given in Fig.5A.

Fig.3 Schematic representation of the two maximal intersections generated by the stratigraphic relationships between taxa a to d.

Fig. 4 A) Cody's original range chart reduced to its maximal intersections (see text). B) Raw UA range chart calculated by UAgraph using the 116 taxa analysed by Cody et al. (2008). The regional reproducibility of the 41 units is given in Fig. 5.

Fig. 5 Correlation between the UA sequences calculated from the Cody et al. database. A) regional reproducibility of the 92 units (see Fig. 2). Black dots = strictly identified UAs. Grey rectangles = coalescences of identified UAs. Column 33 represents the paleomagnetic sequence; B) regional reproducibility of the 41 units (Fig.4b); C) regional reproducibility of Cody et al. range chart applied to the original database (Fig.4a). Column 33 represents the maximal intersections characterizing the original range chart of Cody et al. (loc.cit. Fig. 6). The discontinuities are generated by contradictions between that range chart and the original database. These contradictions are due to the automatic removing of some rare coexistences during the computation of the database.

Fig. 6 Conflicting relationship between two samples and four taxa involved in a cycle of length 4 (Z4 in graph theoretical terminology). In the graph, straight lines (edges) between taxa represent co-occurrences, while arrows (arcs) represent superpositions.

Fig. 7 Conflicting relationship between two samples of the Cody et al. (2008) database computed by UAGraph.

Fig. 8 A) non oriented graph showing the inter-events trivial virtual edges: each edge represents a conflicting relationship of the type  $a>b$ ,  $a<b$ ; B) oriented graph showing the inter-events arcs: each arc belongs to a cycle of length greater than or equal to 3. List of taxa in Online Appendix 5.

Fig 9 Each box represents a single section (1 to 32; the sections with very few data have been omitted) and the local stratigraphic sequences of datums related to the absolute age (2.58 to 3.6 Ma) interpolated by CONOP (original ages in Cody et al. Table 3; list of taxa in Online Appendix 5). The bold horizontal lines represent uncertainties.

Fig. 10 A) non oriented graph showing the inter-FO trivial virtual edges: each edge represents a conflicting relationship of the type  $a>b$ ,  $a<b$ ; B) oriented graph showing the inter-FO arcs: each arc belongs to a cycle of length greater or equal to 3 (list of taxa in Online Appendix 5).

# Figures

Fig 1

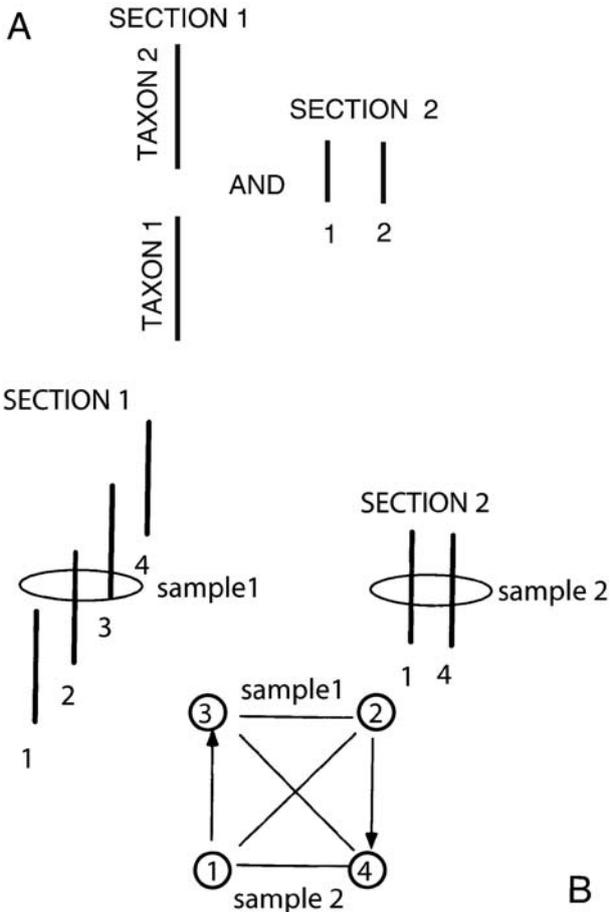








Fig 5

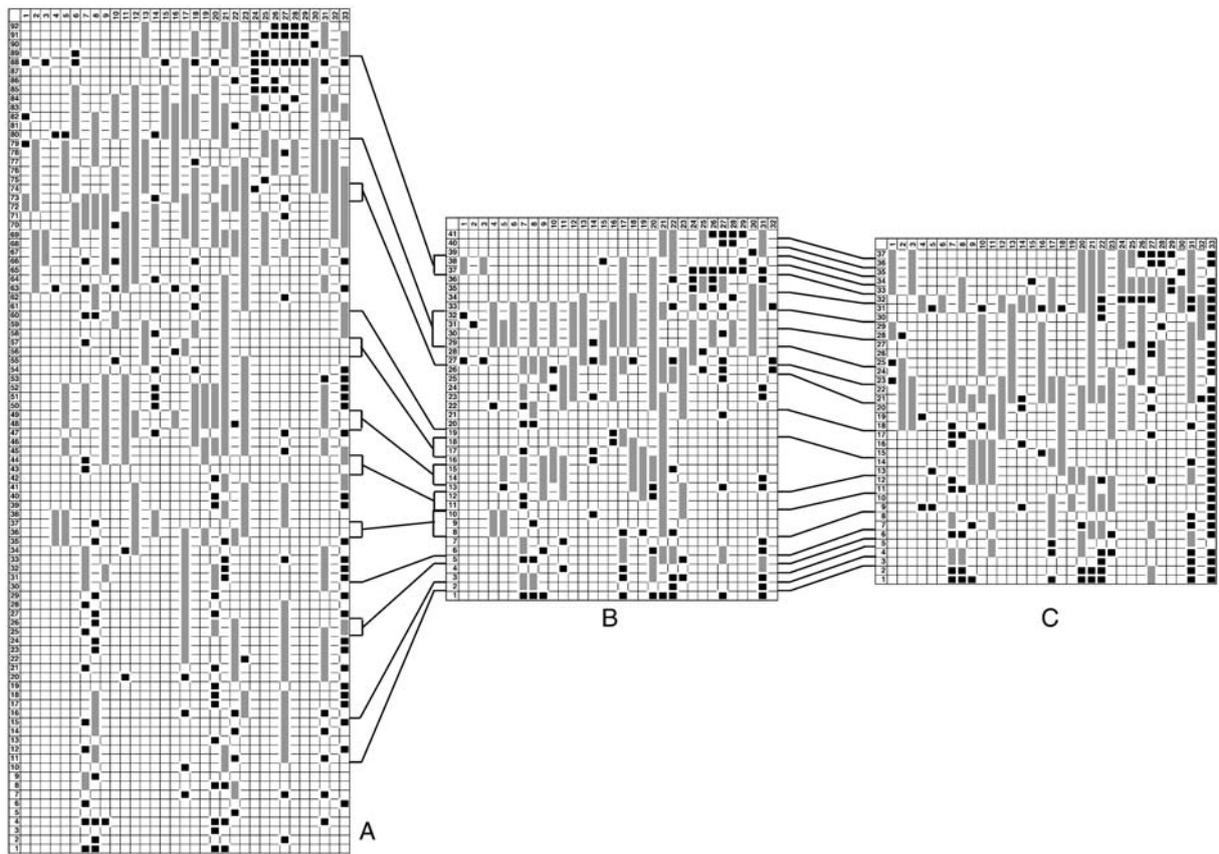


Fig 6

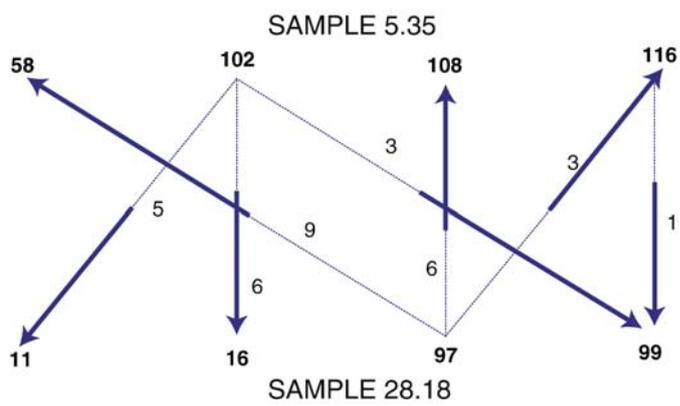


Fig 7

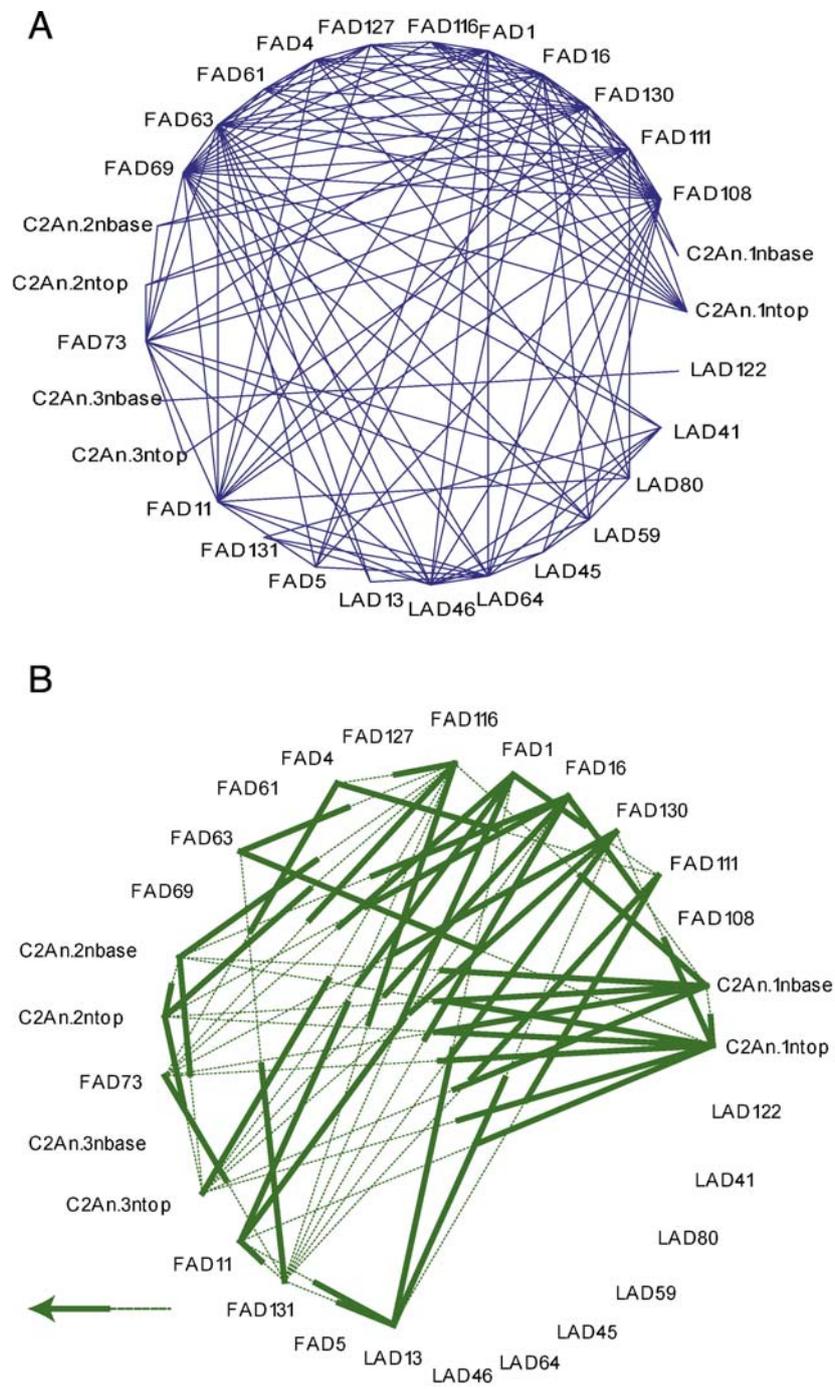


Fig 8

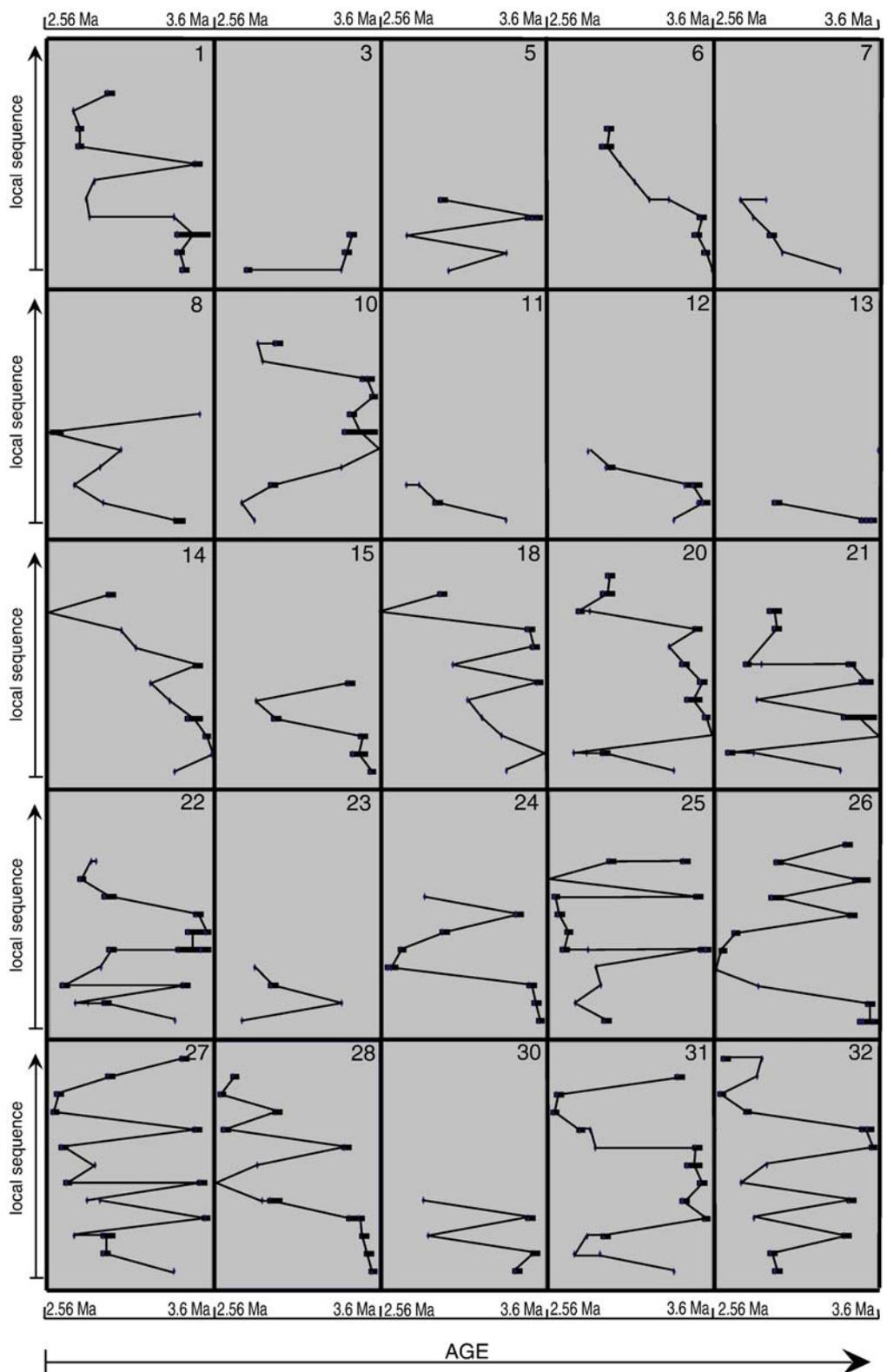


Fig 9

