Monotone Regression – A Method for Combining Dates and Stratigraphy

Irmela HERZOG, Jürgen HANSOHM
The Rhineland Regional Council / The Rhineland Commission for Archaeological Monuments and Sites, Bonn, Germany
University of the Federal Armed Forces, Munich, Germany

Abstract:
The Harris diagram is a well-known means for reconstructing the chronological sequence of archaeological contexts. Floating sequences, i.e. parallel strands in the diagram pose a problem because the chronological sequence of the contexts is not fixed by stratigraphic relationships. Often additional dating information is available, for example dendrochronological evidence, radiocarbon data, or diagnostic finds. Thus, spot dates and date intervals can be assigned to some of the stratigraphic units. This paper presents an algorithm which combines the dates of equal and contemporary contexts and performs plausibility checks for contexts set equal or contemporary. In addition, a stress value is defined which is 0, if no contradiction between the earlier-than-relationships of dated contexts exist. When the stress value exceeds 0, an algorithm known as monotone regression can be used to adjust the dates of the contexts in such a way that they concord with the stratigraphic sequence but are as close as possible to the original date intervals. As an option, the dates for undated contexts may be estimated and periods or phases can be defined on the basis of these dates. This paper will present some results of this method based on simulated data and a dataset from Aegina, Greece. The algorithms described are implemented in version 1.5 of the freeware program *Stratify*.

1 Introduction

The Harris diagram is a well-known means to reconstruct the relative chronological sequence of archaeological contexts (Harris 1989). Floating sequences, i.e. parallel strands in the diagram, pose a problem, because the chronological relationships of two contexts in parallel strands is not determined by stratigraphic observations (Fig. 1). A detailed discussion of the number of possible chronological sequences based on stratigraphic relationships is given by Bibby (2003 and this volume).
An overview of approaches to fixing the floating sequences was presented by Herzog (2006, 7-9). The paper already mentioned that monotone regression is a method to combine absolute dates and the stratigraphic sequence. Monotone regression does not only allow to fix the floating sequences but to assign absolute dates to the contexts. Renfrew and Bahn (1996, 153) point out that the correlation of different dating methods is one of the most promising avenues for future work in chronology. This paper will show that monotone regression is a promising way of walking on the avenue.
The monotone regression approach assumes that the stratigraphic relationships are correct, but that absolute dates which do not agree with the stratigraphic sequence have to be adjusted. The algorithm ensures that inconsistencies between the dates and the stratigraphic relationships are eliminated by changing the dates, but these changes are kept to a minimum. Hansohm (2007) has published a method for calculating the monotone regression, in addition discussing how to deal with date intervals and missing values, finally presenting the analysis of stratigraphic data of the Roman bath in Xanten as an example. Now

the method will become an add-on to the program *Stratify* (www.stratify.org) version 1.5. This paper discusses some refinements of the algorithm and its implementation.

2 Modelling Dates and Date Relationships

Usually, archeological dates are recorded as time intervals rather than points in time. An example discussed in the majority of modern basic introductions to methods in archaeology (e.g. Renfrew/Bahn 1996, 132-133) are radiocarbon dates, with the date intervals given in the form d ± σ, where d is the mean of the Gaussian probability distribution and σ the standard deviation. This means that the chance that the date lies within the interval [d-σ, d+σ] is approximately 68% and /even/ as high as 95% when considering the interval [d-2σ, d+2σ].

What is probably the most popular dating method in archaeology is based on pottery typology (or some other find typology). Models for the production of a pottery style assume that the initial production is low, gradually increases to a peak, and fades away as another pottery style becomes popular. Due to their shape, these type production curves are often called battleship curves. Seriation is a method for sorting assemblages which is based on the model of battleship type production curves (e.g. Renfrew/Bahn 1996, 116-118). Such a pottery production curve bears some similarity with the bell-shaped curve of the Gaussian probability distribution, and therefore, it seems appropriate to treat dates derived from pottery typology and calibrated radiocarbon dates in the same way.

In the cases of radiocarbon and pottery production dates, date intervals are an expression of uncertainty. But time intervals may also represent duration, and some deposits have accumulated over a long period of time without any recognizable change in the structure of the deposit. In a presentation given at the CAA conference in 2000, Holst (2001) proposed an extension of the Harris diagram methodology by introducing time spans for the contexts and modeling 15 types of temporal relationships between two contexts. In 2004, he was able to present a computer program which allowed creating a relative chronological sequence based on these relationships, and this is displayed as a graph similar to the Harris diagram (Holst 2004a). The concept of Holst is closer to archaeological reality than the simple stratigraphic relationships introduced by Harris, and according to Holst, these simple relationships represent a generalisation which largely disregards the ambiguities of archaeological data. Holst's methodological framework also introduces additional possibilities of expressing uncertainty. Given that the model proposed by Holst is closer to archaeological reality and a computer program is available to deal with this kind of data (Holst 2004b), then why is the archaeological community so reluctant to adopt this model? Holst gives an example which sheds some light on this issue: He tries to establish temporal relationships between features in a settlement, these features include longhouses, fences, and farms. Quite a few stratigraphic units are grouped to form each of these features, and in our view it is more appropriate to deal with the basic individual stratigraphic units and their relationships. Most stratigraphic units were created within a small time span, whereas the creation time spans for features are longer. For this reason, representing the time span of a stratigraphic unit as a single date is in most cases sufficiently close to reality. Based on this simplification, the analysis of the stratigraphy is much easier than dealing with 15 types of relationships. If the excavator knows that a stratigraphic unit has accumulated over a long period of time, it is recommended in this special case to create two boxes in the Harris diagram, the start event box and the end event box. This is easier than to create a model providing two dates, the start and end dates, for each stratigraphic unit. Therefore, we decided on treating the date intervals as intervals expressing uncertainty rather than intervals indicating duration, and we focus on calculating just one date for each stratigraphic unit. The traditional stratigraphic relationship model is adequate to describe all the temporal relationships between two stratigraphic units each of which is modeled as a point in time. We are aware that Doerr (in print) is right in pointing out that all models treating events as points in time are inherently inconsistent with reality and not observable, but we argue that the model implicitly used in the Harris diagram is a successful simplification.

3 The *Stratify* Implementation of Monotone Regression

The *Stratify* implementation of monotone regression will be explained on the basis of the small stratigraphic dataset presented in Fig. 1. Twenty contexts were simulated; the oldest context is F01, while the most recent context is F20. Each context was assigned a creation year and then typical date intervals for these creation years were chosen manually, excepting four contexts where the creation years were erased (Tab. 1). Note that errors were introduced on purpose; for two contexts (F09, F16) the simulated date is not within the date interval. F13 and F14 were set equal, F14 and F15 were set contemporary as well as F10 and F12.

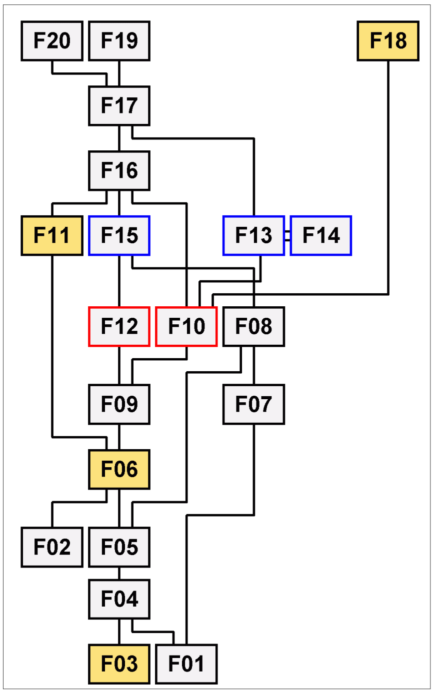

Fig. 1 - Simulated dataset consisting of 20 contexts. According to the diagram, F07 is contemporary with F09. The stratigraphic relationships do not rule out that F07 is younger than F09 or older than F04. The contexts marked by yellow backgrounds are undated. The two contexts marked by red frames form a tie group, as do the three contexts with blue frames.

Context	Simulated year	Interval min.	Interval max.	Interval mean	Weight	Monotone regression result
F01	11	-50	100	25	0.081	25
F02	56	50	100	75	0.140	75
F03	81					25
F04	116	0	200	100	0.071	100

Context	Simulated year	Interval min.	Interval max.	Interval mean	Weight	Monotone regression result
F05	155	125	175	150	0.140	150
F06	157					162
F07	184	150	300	225	0.081	193
F08	199	150	200	175	0.140	193
F09	205	150	200	175	0.140	175
F10	**225**	**150**	**350**	**250**	**0.071**	**290**
F11	259					283
F12	**299**	**275**	**325**	**300**	**0.140**	**290**
F13	**301**	**300**	**375**	**337**	**0.115**	**364**
F14	**309**	**275**	**425**	**350**	**0.081**	**364**
F15	**364**	**364**	**364**	**364**	**1.000**	**364**
F16	392	400	450	425	0.140	417
F17	430	300	500	400	0.071	417
F18	446					500
F19	470	400	550	475	0.081	475
F20	500	450	550	500	0.100	500

Tab. 1 - Dates assigned to the simulated dataset depicted in Fig. 1. The two contexts marked in red form a tie group as do the three blue contexts.

The simulated dates of the contexts are not taken into account when calculating the monotone regression result but will be used to evaluate the accuracy of the result. In a first step, the mean of the date interval is calculated which is considered to be an estimate of the true date of each context with a date interval. In addition, the weight of the date is defined so that it indicates the reliability of the date estimation. The reliability of contexts with a large date interval is low. Hansohm (2007, 1049) suggests to set the weight to the inverse of the length of the date interval increased by 1, so that the weight of F20 is 1/(550-450+1). In Tab. 1 another weight selection method was chosen, i.e. the square root of the weights defined by Hansohm. Both options are implemented in *Stratify*.

The algorithm consists of the following steps:

- First, the date interval means and the weights are calculated.
- Then the dates of equal and contemporary contexts are combined, i.e. two contexts which are equal or contemporary are assigned the same date.
- A list of all pairs of dated contexts with an earlier-than-relationship is created.
- The total stress value is calculated: Each pair of dated contexts where the dates are not consistent with the earlier-than-relationship contributes to the total stress value.
- Monotone regression is used to estimate new date values that are as close as possible to the interval mean dates but which are congruent with the stratigraphic relationships.

- Optionally, dates for contexts without initial date are estimated.
- On the basis of these dates, periods or phases may be defined.

Details of these steps will be described below.

3.1 Combining the dates of equal and contemporary contexts

If contexts are set equal or contemporary, they are said to form a tie group. The example dataset includes two tie groups (Tab. 1). The dates of the contexts belonging to a tie group must be identical. Hansohm coined the term tie group, and he calculates the date of a tie group as the weighted average of the interval means, while the weight of the tie group is the sum total of the members' weights. In archaeology, a well-known example of combining the dates of tie groups is the radiocarbon combination of dates (Baxter 2003, 189-190).

Let d_1, d_2, ..d_n be the dates of n contexts in a tie group and w_1, w_2, .., w_n the corresponding weights, then the two alternative ways of date combination are given by the following formulas:

(1) $d_t = (w_1 d_1 + w_2 d_2 + \dots + w_n d_n) / (w_1 + w_2 + \dots + w_n)$

(2) $w_t = w_1 + w_2 + \dots + w_n$

(3) $d_t = (w_1^2 d_1 + w_2^2 d_2 + \dots + w_n^2 d_n) / (w_1^2 + w_2^2 + \dots + w_n^2)$

(4) $w_t = \sqrt{(w_1^2 + w_2^2 + \dots + w_n^2)}$

where formulas (1) and (2) are Hansohm's date calculations, and (3) and (4) are the corresponding formulas for radiocarbon date combination. The value d_t is the date of the tie group, w_t is the weight. When combining four dates with equal weight, the resulting weight of the Hansohm method is four times the individual weights, whereas with radiocarbon combination the new weight of the tie group is twice the original weight. While both date combination methods have been implemented in *Stratify* 1.5, the radiocarbon date combination was chosen for the example presented here. However, the Hansohm formulas are in close agreement with the theory of monotone regression, and it might be a line of further research to modify the method of monotone regression in such a way that whenever date combinations are necessary the radiocarbon date combination formulas are used.

The model of the Gaussian normal distribution can be used to test the plausibility of an equal or contemporary relationship. This is done by calculating the difference between the two dates assigned to the contexts involved in the relationship, and if the value exceeds certain limits, two different warnings are produced similar to yellow and red traffic lights: If the difference is greater than the sum total of the standard deviations, a yellow warning is issued, whereas a red warning is shown if this difference exceeds twice this sum (Fig. 2).

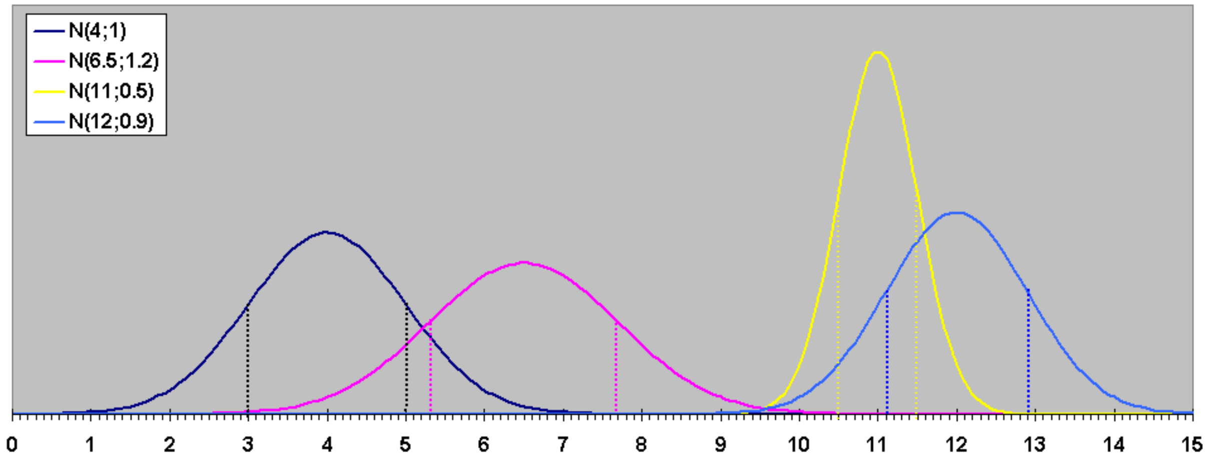

Fig. 2 - Plausibility check for combining two dates with Gaussian normal distribution. When trying to combine the dark blue and the pink dates, a yellow warning is issued because the one sigma intervals marked by dotted lines do not overlap. No warning results when combining the two dates corresponding to the yellow and light blue curves, because the one sigma intervals overlap. Since the two sigma intervals of the pink and the yellow curves do not overlap, a red warning will be shown if two contexts with these two date distributions are within one tie group.

Stratify lists the warnings in the optional protocol file which is displayed at the end of the date calculation process. For the test dataset, no warnings were issued as the date intervals of the contexts set equal or contemporary overlap.

3.2 Setting up the list of relationships

The next step of the algorithm is to construct a list of earlier-than relationships between pairs of dated contexts. Each tie group is treated like a single context, i.e. all earlier-than relationships of these contexts are assigned to one representative of the tie group. It is checked that the relationships do not form cycles, and redundant relationships are removed (for explanations of the terminology used here see the help file of *Stratify*). The relationship list for the test dataset comprises the following relationships:

F01 < F04, F01 < F07
F02 < F09, F02 < F16
F04 < F05
F05 < F08, F05 < F09, F05 < F16
F07 < F08
F08 < F13/F14/F15
F09 < F10/F12
F10/F12 < F13/F14/F15
F13/F14/F15 < F16
F16 < F17
F17 < F19, F17 < F20

Note that the relationship F02 < F06 is omitted because F06 is without date. Instead, the indirect relationships F02 < F09 and F02 < F16 are included in the list.

3.3 Stress calculation

The list of relationships is parsed, and whenever the date of the earlier context exceeds the date of the later context, the stress introduced by this relationship is calculated: The stress contribution of this relationship pair is the date difference multiplied by the sum total of the weights of the two contexts and divided by the total of all relationship pair weight sums. All the stress contributions are added up to result in a total stress value.

If the total stress is 0, no contradictions between the stratigraphic relationships and the absolute dates have been detected, hence monotone regression is not needed in this case. In the test dataset, only two relationship pairs contribute to the total stress value: F07 < F08 and F16 < F17. The estimated date of F07 (225) exceeds the date of F08 (175). The relationship F07 < F08 contributes 1.72 to the total stress value, and the contribution of F16 < F17 is 0.82, which means that the total stress value is 2.54.

On selecting the Detailed Report option, a list of earlier-than relationships which contribute to the stress value is created. It is recommended to double-check all relationships that make an exceptionally large contribution to the total stress value, because gross errors may distort the results of the monotone regression.

3.4 The core of monotone regression

The aim of monotone regression is to find date estimates which are as close as possible to the initial absolute dates but which do ensure that the total stress is 0. The sum total of differences between initial absolute dates and the date estimates is minimised in a least squares sense. Another method using the least squares approach is linear regression, and this method is well known among archaeologists interested in statistics (e.g. Shennan 1997, 133-139). Linear regression assumes that a linear relationship exists between the variable X and the dependent variable Y while some noise has been added to the observations of variable Y. Least squares methods are based on the assumption that the noise is distributed normally. It is therefore important to detect and eliminate gross errors before starting a least squares optimisation procedure.

Since according to Hansohm (2007), the direct solution of the monotone regression problem in most cases involves a large number of calculations, he proposes an iterative solution. The iterative algorithm decreases the stress value step by step, i. e. the larger the number of iterations, the closer the iteration result is to the true solution. For this reason, a spiral was chosen as a symbol for the algorithm since the spiral curve is progressively approaching the central point.

For the simple test dataset, the algorithm finishes after only two iterations. The dates of F07, F08, F16, and F17 are adjusted. The two contexts F07 and F08 are set to the same date, which is closer to the initial date of F08, as the weight of F08 is greater than that of F07.

Once the algorithm has assigned new dates to the contexts, *Stratify* checks if the new dates are within the initial date intervals. If this is not the case, an appropriate message is added to the optional protocol file.

3.5 Date estimation for contexts without date

Hansohm (2007) suggests including contexts without initial date in the computation by choosing a large interval or a very low weight for these contexts. In the example of the Roman bath of Xanten, one million iterations were necessary to get sufficiently close to the true solution. When the dataset is restricted to the dated contexts and their relationships, the algorithm stops after 84 to 135 iterations, depending on the choice of weights and the date combination method used.

Another approach to deal with missing dates has been implemented in the *Stratify* program. First, new dates are assigned to all dated contexts according to the monotone regression method. Then the missing dates for contexts without date are estimated depending on the class of the context. There are three classes of contexts without a date:

(1) Contexts without any dated predecessor like F03 in the test dataset.
These contexts are assigned the minimum date of all dates in the dataset. The smallest estimated date is the date of F01 (25), therefore the date of F03 is set to 25.

(2) Contexts without any dated successor like F18 in the test dataset.
The dates of these contexts are set to the maximum date of all dates in the dataset. For this reason, the F18 was assigned a date of 500.

(3) Contexts with dated predecessors and dated successors like F06 and F11 in the test dataset.
For these contexts the minimum date of the dated successors and the maximum date of the dated predecessors is determined. The undated contexts are assigned the average of these two dates. The following example presents the calculation of the date of F06:

D1 = min(date(F09), date(F16)) ! min(175, 417)
D2 = max(date(F02), date(F05)) ! max(75, 150)
date(F06) = (D1 + D2)/2 ! (150+175)/2

Only if the undated contexts are in a chain of relationships, i.e. only if one dated predecessor and one dated successor exist and all the contexts along the chain connecting the dated predecessor and the dated successor have exactly one earlier-than and one later-than relationship, then the dates will be distributed regularly along the chain.

3.6 Phase or period assignment

When all the contexts are dated and the context dates are consistent with the stratigraphic relationships, the phase or period assignment is straightforward. The user selects the number of phases or periods to be created, the phase or period names, and the date intervals assigned to these periods or phases. If a context's date falls within a phase's date interval, the context's phase is set to the appropriate phase name. Fig. 3 shows the *Stratify* user interface for setting the phase names and the start and end years.

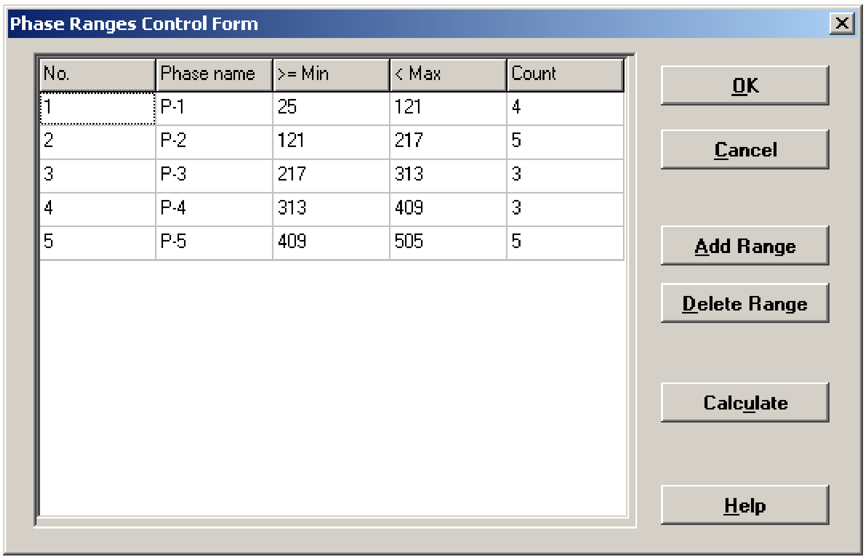

Fig. 3 - *Stratify* interface used for phase or period assignment. The “Count” column is read-only and is updated by pressing the ”Calculate” button.

4 Results

For the test dataset, the mean difference between the initial simulated date and the date calculated by monotone regression with Hansohm weights and Hansohm date combination is 23.8 years. Since the average date interval length for the dated contexts exceeds 100 years, this result is quite good. The correlation between the simulated dates and the monotone regression dates is 0.98. With the alternative weight calculation and radiocarbon combination of dates, the correlation coefficient is approximately the same, and the mean difference between simulated and calculated dates is slightly lower (23.7 years).

5 Monotone Regression of the Kolonna Dataset

The Kolonna dataset from Aegina, Greece (Gauß, Smetana 2007), serves as a practical example. It consists of 336 stratigraphic units, with dates ranging from Early Helladic II (about 2650 BC – 2200 BC) to modern. The focus of the excavation was on Helladic contexts, but according to the provisional dating by the excavator, 27 modern and 40 intermediate contexts were recorded as well, where the term intermediate refers to all periods between Helladic and modern, i.e. this is the time span beginning roughly at 1190 BC and ending at 1492 AD. Even at the beginning of this excavation, Helladic features were visible on the surface. Therefore, the Harris diagram constructed from stratigraphic relationships only shows contexts on the first level with dates varying within a time interval of more than 3000 years. This is one of the reasons why additional dating information has to be included in the layout of the Kolonna Harris diagram. Date intervals were assigned to 297 of the contexts; these date ranges were mainly estimated on the basis of the finds and the type of features excavated. Eleven time intervals were determined by the radiocarbon method. One of the aims of the monotone regression analysis in this case is to check if these radiocarbon dates are in agreement with the dates resulting from archaeological typology.

The monotone regression procedure was run including all available dating information. The algorithm first identified 27 tie groups, with the largest tie group consisting of 17 contexts. Two of the contexts belonging to the largest tie group have a radiocarbon date. When combining the dates of these two contexts by the radiocarbon method, the result is 1847 BC; when taking all 17 contexts into account, a date of 1850 BC is calculated. The dates of the contexts within each tie group were estimated by the radiocarbon date combination method. All the dates to be combined were plausible, i.e. none of the dates of any tie group member led to a warning.
The relationship list that was set up contained 268 entries. Twelve of these contributed to the total stress value of 2.75. The stress relationships involved five out of the eleven contexts with radiocarbon date. The stress contributed by the relationships with radiocarbon dated contexts is 92% of the total stress value (2.53). Some of the problems might be due to the fact that the average date interval size for the radiocarbon dates exceeds the corresponding value of the archaeologically dated contexts (198 versus 119 years). In addition, the archaeologist, trying to avoid chronological inconsistencies, might have had the stratigraphic sequence in mind while dating the contexts.
The monotone regression algorithm finished after 54 iterations. Afterwards two contexts were identified whose new dates were not within the initial date intervals. Both contexts had been dated by the radiocarbon method. In addition, two relationships of these contexts contributed most to the total stress value (individual contributions: 0.94 and 0.39). For this reason, it might be a good idea to re-check these radiocarbon dates.
In another experiment performed with a copy of the Kolonna dataset, the radiocarbon dates were deleted and monotone regression run again. The initial stress value was much lower in this case (0.22), the algorithm already stopped after seven iterations. New dates were estimated for the contexts whose radiocarbon dates had been deleted and these new dates were compared with the radiocarbon dates. In the case of the two contexts belonging to the tie group consisting of 17 members, the date estimates were very close to the radiocarbon date (difference: 15 years). Only for two other contexts, the differences were below 100 years, while for five contexts the difference was in the range of 200 to 315 years. But three out of these five contexts were at the bottom of the Harris diagram, i.e. no earlier contexts had been recorded, therefore the contexts were assigned the minimum year of all the dated contexts, which is obviously not appropriate in this situation. Having explained three of the large differences, the two other contexts with large differences are the well-known bad ones, i.e. the two contexts with the largest contribution to the stress value in the analysis including the radiocarbon dates.
These two experiments with the Kolonna data show how monotone regression can be applied to identify problematic dates in a stratigraphic dataset.

6 Acknowledgements

We would like to thank Walter Gauß for providing the Kolonna dataset along with detailed information concerning the data. The radiocarbon dates of this dataset were determined by VERA, a centre for accelerator mass spectrometry (AMS) at the University of Vienna, Institut für Isotopenforschung und Kernphysik in Vienna. The VERA institute will present the radiocarbon dates of Kolonna in a separate publication. We would like to record our gratitude to the team of the 12th Vienna Workshop for perfectly organising an inspiring conference. Particular thanks are also due to Günter Merboth for improving the English of this paper.

References

Baxter, M., 2003. Statistics in Archaeology (London, New York), 189-190.

Bibby, D. 2003. Permutations of the Multilinear Stratigraphic Sequence: Nature, Mathematics & Consequences. Paper presented at the Workshop 7 “Archäologie und Computer”, Vienna. Powerpoint file on CD (Vienna 2003).

Bibby, D. (this volume). What does the Harris Matrix mean? Some thoughts on the multilinearity of the stratigraphic sequence. Paper presented at the Workshop 12 “Archäologie und Computer”, Vienna.

Doerr, M. (in print). Supporting Chronological Reasoning in Archaeology. In: Proceedings of the 32nd Conference on Computer Applications and Quantitative Methods in Archaeology CAA2004, Prato, Italy, April 2004.

Gauß, W., Smetana, R. 2007. Aegina Kolonna, The Ceramic Sequence of the SCIEM 2000 Project. In: F. Felten (ed.) Ägina - Kolonna. Forschungen und Ergebnisse. Band I / F. Felten, W. Gauss, R. Smetana, Middle Helladic Pottery and Synchronisms. Proceedings of the International Workshop held at Salzburg October 31st – November 2nd, 2004. Contributions to the Chronology of the Eastern Mediterranean, Vol. XIV (Vienna 2007): 57-80.

Hansohm, J. 2007, Algorithms and error estimations for monotone regression on partially preordered sets. Journal of Multivariate Analysis 98: 1043 – 1050.

Harris, E. C. 1989. Principles of archaeological stratigraphy, Academic Press, London.

Herzog, I. 2006. No News from Stratigraphic Computing? Paper presented at the Workshop 10 “Archäologie und Computer. Kulturelles Erbe und Neue Technologien”, November 7-10, 2005. Vienna. PDF file on CD (Vienna 2006).

Holst, M. K. 2001. Formalizing Fact and Fiction in Four Dimensions: A Relational Description of Temporal Structures in Settlements. In: Stančič, Z., Veljanovski, T. (Eds.), Computing Archaeology for Understanding the Past, CAA 2000, BAR Int. Ser. 931, pp. 159-163.

Holst, M. K. 2004a. Complicated Relations and Blind Dating: Formal Analysis of Relative Chronological Structures. In: Buck, C. E., Millard, A. R. (eds.) Tools for Constructing Chronologies. Lecture Notes in Statistics 177 (Springer, London/Berlin/Heidelberg): 129-147.

Holst, M. K. 2004b: Tempo. http://www.math.ku.dk/~holst/tempo/ [29 Dec 2007]

Renfrew, C./Bahn, P. 1996. Archaeology. Theories, Methods and Practice (London 1996).

Shennan, St. 1997. Quantifying Archaeology (Edinburgh 1997).

Stratify: www.stratify.org