



Precision of the current methods to measure the alkenone proxy $U_{37}^{K'}$ and absolute alkenone abundance in sediments: Results of an interlaboratory comparison study

A. Rosell-Melé

Department of Geography, University of Durham, Durham, DH1 3LE, England, UK (antoni.rosell@durham.ac.uk)

E. Bard

Centre Européen de Recherche et d'Enseignement de Géosciences de l'Environnement, Université d'Aix-Marseille III, Europole de l'Arbois, BP80, F-13545 Aix-en-Provence cedex 4, France

K.-C. Emeis

Institut für Ostseeforschung Warnemünde, Seestrasse 15, D-18119 Warnemünde, Germany

J. O. Grimalt

Departamento Química Ambiental, CID-CSIC, Jordi Girona 18, Barcelona E-08034, Catalonia, Spain

P. Müller and R. Schneider

Fachbereich Geowissenschaften, Universität Bremen, FB 5 - Geowissenschaften, Postfach 330 440, D-28334 Bremen, Germany

I. Bouloubassi

Laboratoire de Physique et Chimie Marines, Université Pierre et Marie Curie, Case 134, Tour 25, 5ème étage, 4 place Jussieu, F-75252 Paris Cedex 05, France

B. Epstein

Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882-1197, USA

K. Fahl

Alfred-Wegener-Institute for Polar and Marine Research, Columbusstrasse, D-27568 Bremerhaven, Germany

A. Fluegge

Alfred-Wegener-Inst. for Polar and Marine Research, Am Handelshafen 12, D-27570, Bremerhaven, Germany

K. Freeman

Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA

M. Goñi

Department of Geological Sciences, University of South Carolina, Columbia, South Carolina 29208, USA

U. Güntner

Institut für Chemie und Biologie des Meeres, University of Oldenburg, P.O. Box 2503, D-26111 Oldenburg, Germany

D. Hartz

Department of Geological Sciences, University of South Carolina, Columbia, South Carolina 29208, USA

S. Hellebust

Department Fossil Fuels and Environmental Geochemistry, University of Newcastle, Newcastle upon Tyne NE1 7RU, England, UK

T. Herbert

Department of Geological Sciences, Box 1846, Brown University, Providence Rhode Island 02912, USA

M. Ikehara

Institute Low Temperature Science, Hokkaido University, N19 W8, Kita-ku, Sapporo 060, Hokkaido, Japan

R. Ishiwatari

Department of Chemistry, Faculty of Science, Tokyo Metropolitan University, Minami-Ohsawa 1-1, Hachoji, Tokyo, 192-03, Japan

K. Kawamura

Institute Low Temperature Science, Hokkaido University, N19 W8, Kita-ku, Sapporo 060, Hokkaido, Japan

F. Kenig

Department of Geological Sciences, 845 W Taylor Street, University of Illinois at Chicago, Chicago, Illinois 60607-7059, USA

J. de Leeuw

Department of Organic Geochemistry and Toxicology, Netherlands Institute for Sea Research, P.O. Box 59,1790 AB, Den Burg, Texel, Netherlands

S. Lehman

Institute of Arctic and Alpine Research, Campus Box 450, University of Colorado, Boulder Colorado 80309, USA

L. Mejean

Laboratoire de Physique et Chimie Marines, Université Pierre et Marie Curie, Case 134, Tour 25, 5ème étage, 4 place Jussieu, F-75252 Paris Cedex 05, France

N. Ohkouchi

Institute Low Temperature Science, Hokkaido University, N19 W8, Kita-ku, Sapporo 060, Hokkaido, Japan

R. D. Pancost

Department of Geosciences, Pennsylvania State University, University Park, Pennsylvania 16802, USA

C. Pelejero

Departamento Química Ambiental, CID-CSIC, Jordi Girona 18, Barcelona E-08034, Catalonia, Spain

F. Prahl

College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA

J. Quinn

Graduate School of Oceanography, University of Rhode Island, Narragansett, Rhode Island 02882-1197, USA



J.-F. Rontani

Laboratoire d'Océanographie et de Biogéochimie (UMR 6535), Centre d'Océanologie de Marseille (OSU), Faculté des Sciences de Luminy - case 901, F-13288 Marseille, France

F. Rostek

Centre Européen de Recherche et d'Enseignement de Géosciences de l'Environnement, Université d'Aix-Marseille III, Europole de l'Arbois, BP80, F-13545 Aix-en-Provence cedex 4, France

J. Rullkötter

Institut für Chemie und Biologie des Meeres, University of Oldenburg, P.O. Box 2503, D-26111 Oldenburg, Germany

J. Sachs

Institute of Arctic and Alpine Research, Campus Box 450, University of Colorado, Boulder Colorado 80309, USA

T. Blanz

Institut für Meereskunde - Meereschemie, Duesternbrooker Weg 20, D-24105 Kiel, Germany

K. Sawada

Department of Chemistry, University of Tsukuba, Tsukuba, Ibaraki 305-8571, Japan

D. Schulz-Bull

Institut für Meereskunde - Meereschemie, Duesternbrooker Weg 20, D-24105 Kiel, Germany

E. Sikes

School of Environmental and Marine Sciences, University of Auckland, Private Bag 92019, Auckland, New Zealand

C. Sonzogni

Centre Européen de Recherche et d'Enseignement de Géosciences de l'Environnement, Université d'Aix-Marseille III, Europole de l'Arbois, BP80, F-13545 Aix-en-Provence cedex 4, France

Y. Ternois

Institute Low Temperature Science, Hokkaido University, N19 W8, Kita-ku, Sapporo 060, Hokkaido, Japan

G. Versteegh

Department of Organic Geochemistry and Toxicology, Netherlands Institute for Sea Research, P.O. Box 59, 1790 AB, Den Burg, Texel, Netherlands

J. K. Volkman

CSIRO Marine Research, GPO Box 1538, Hobart, Tasmania 7001, Australia

S. Wakeham

Skidaway Institute of Oceanography, 10 Ocean Science Circle, Savannah, Georgia 31411, USA

[1] **Abstract:** Measurements of the $U_{37}^{K'}$ index and the absolute abundance of alkenones in marine sediments are increasingly used in paleoceanographic research as proxies of past sea surface temperature and haptophyte (mainly coccolith-bearing species) primary productivity, respectively. An important aspect of these studies is to be able to compare reliably data obtained by different laboratories from a wide variety of locations. Hence the intercomparability of data produced by the research community is essential. Here we report results from an anonymous interlaboratory comparison study involving 24 of the leading laboratories that carry out alkenone measurements worldwide. The majority of laboratories produce data that are intercomparable within the considered confidence limits. For the



measurement of alkenone concentrations, however, there are systematic biases between laboratories, which might be related to the techniques employed to quantify the components. The maximum difference between any two laboratories for any two single measurements of $U_{37}^{K'}$ in sediments is estimated, with a probability of 95%, to be $<2.1^{\circ}\text{C}$. In addition, the overall within-laboratory precision for the $U_{37}^{K'}$ temperature estimates is estimated to be $<1.6^{\circ}\text{C}$ (95% probability). Similarly, from the analyses of alkenone concentrations the interlaboratory reproducibility is estimated at 32%, and the repeatability is estimated at 24%. The former is compared to a theoretical estimate of reproducibility and found to be excessively high. Hence there is certainly scope and a demonstrable need to improve reproducibility and repeatability of $U_{37}^{K'}$ and especially alkenone quantification data across the community of scientists involved in alkenone research.

Keywords: alkenones; $U_{37}^{K'}$; interlaboratory comparison.

Index terms: Organic geochemistry; instruments and techniques; paleoceanography; organic marine chemistry.

Received January 22, 2001; **Revised** April 5, 2001; **Accepted** April 19, 2001; **Published** July 6, 2001.

Rosell-Melé A., et al., 2001. Precision of the current methods to measure the alkenone proxy $U_{37}^{K'}$ and absolute alkenone abundance in sediments: Results of an interlaboratory comparison study, *Geochem. Geophys. Geosyst.*, vol. 2, Paper number 2000GC000141 [9416 words, 13 figures, 6 tables]. Published July 6, 2001.

Theme: Alkenones **Guest Editor:** John Hayes

1. Introduction

[2] A significant breakthrough in paleoenvironmental sciences has proved to be the discovery in the mid-1980s that the marine sedimentary record of certain biomarkers could provide insights into past changes in sea surface temperature (SST) [Brassell *et al.*, 1986]. Several molecular indices have been proposed as proxies for SST, chiefly on the basis of the quantification of the relative abundance of diunsaturated and triunsaturated C_{37} alkenones in marine sediments, of which the most widely applied is the $U_{37}^{K'}$ index [Prah and Wakeham, 1987]. The $U_{37}^{K'}$ index has been steadily gaining acceptance for paleotemperature reconstruction, and an increasing number of research groups worldwide are now measuring it routinely using in-house facilities. Some questions, however, still need to be addressed as to date, only relatively few

studies have focused on the analytical constraints of measuring $U_{37}^{K'}$ [Rosell-Melé, 1994; Rosell-Melé *et al.*, 1995; Villanueva, 1996; Villanueva and Grimalt, 1996; Sonzogni *et al.*, 1997; Ternois *et al.*, 1997; Villanueva *et al.*, 1997a; Comes and Rosell-Melé, 1999]. For instance, is there any guarantee that $U_{37}^{K'}$ -SST estimates produced by any one laboratory can be reproduced elsewhere with an error smaller than $1-2^{\circ}\text{C}$? This question is important when one considers that the temperature difference between the Holocene and the Last Glacial Maximum in many areas of the world may not have exceeded 2°C [e.g., CLIMAP Project Members, 1981]. In addition, there is no standard against which the accuracy of $U_{37}^{K'}$ data can be compared, or the deviation in the results between laboratories ascertained. Uncertainties are potentially important in view of differences in the analytical procedures used at various stages of $U_{37}^{K'}$ determination



(e.g., extraction, cleanup, solvent extract fractionation) by those groups involved in alkenone studies.

[3] Similarly, the absolute abundance of C_{37} alkenones in sediments is used to derive estimates of paleoproductivity of the alkenone producers (see discussion by *Prahl and Muehlhausen* [1989] and *Brassell* [1993]). This measurement is gaining in recognition as a useful proxy in studies of oceanographic variability during the Quaternary. Thus it is important to ascertain whether the quantification procedures used by different laboratories yield values that are comparable or at least that errors and differences between laboratories are not larger than natural variations in the proxies at the different scales of time and space that may be considered.

[4] To address these issues, a number of laboratories worldwide have participated in an anonymous interlaboratory comparison study (also called round-robin, intercalibration exercise, or laboratory performance study according to the nomenclature of the International Union of Pure and Applied Chemistry (IUPAC) [*Horwitz*, 1994]) and have analyzed the alkenone composition of standard mixtures and of several sediment samples. A collaborative study of this kind is an exacting method for assessing the capability of a laboratory (or rather their analysts) and its analytical protocols. One objective of this study has been to evaluate whether there are significant differences among the laboratories' results (for $U_{37}^{K'}$ and C_{37} alkenone absolute quantification). A secondary objective has been to evaluate if there are any significant differences related to the analytical methods that are used. Finally, we have investigated if the overall repeatability and reproducibility of the methods should be improved (i.e., are the current

methods to measure alkenone relative and absolute abundance sufficiently precise?). In this study, there has not been an a priori selection of methods, nor is the aim to produce a ranking of laboratories or to provide a form of accreditation.

2. Experimental Section

[5] Each laboratory was provided with eight test samples (five sediments and three mixtures of alkenone standards) and an identification code. Participants were instructed to include the test samples as part of a batch of samples that were currently being analyzed to make the results representative of their routine procedures. They were also requested to carry out as many replicate analyses as possible (six preferably, with three being the recommended minimum), so that any subsequent statistical treatment was meaningful. Materials were sent to 30 laboratories and 24 returned results by the set deadline.

2.1. Standards

[6] The provided alkenone standards [*Rechka and Maxwell*, 1988] were mixed in three different proportions. Aliquots were taken from two parent standard solutions ($C_{37:3}$ alkenone: 43.4 $\mu\text{g/mL}$; $C_{37:2}$ alkenone 38.4 $\mu\text{g/mL}$) to prepare mixtures in approximately the proportions 50/50, 25/75, and 75/25 (i.e., $U_{37}^{K'} = 0.4695$, 0.726, and 0.228, respectively, calculated from the concentration of the solutions). Note that the relative molar responses (RMR) for the alkenones can be expected to be very similar on the commonly employed flame ionization detector. Thus, theoretically, these are $\text{RMR } C_{37:3} = 3534$ and $\text{RMR } C_{37:2} = 3556$ and the $U_{37}^{K'}$ derived for the standard mixtures using response factors would be 0.470, 0.726, and 0.228 for the mixtures 50/50, 25/75, and 75/25, respectively (RMR calculated using values for functional groups

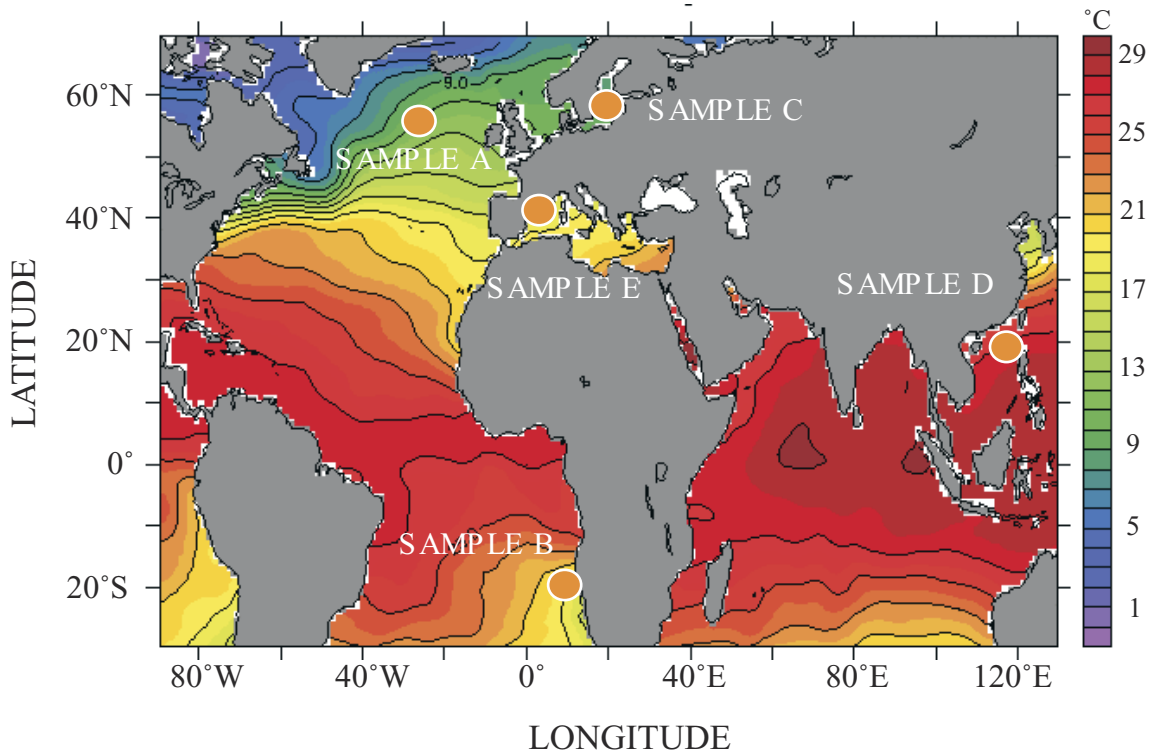


Figure 1. Map of sea surface temperatures (annual mean at 0 m [Levitus and Boyer, 1994]) showing the location of the sediment samples. Samples were provided by E. Bard (sediment D; representative of pelagic conditions), J. Grimalt (sediment E; representative of estuarine/coastal conditions), K.-C. Emeis (sediment C; representative of lacustrine conditions), P. Müller (sediment B; representative of pelagic conditions), and A. Rosell-Melé (sediment A; representative of pelagic conditions).

by Dabrio *et al.* [1971] (after the work of Ackman [1964]). From each solution, 100 μL was added to a vial, taken to dryness, capped, and sent to each participant. The participants were asked to dissolve each standard in 100 μL of any solvent they chose and to inject several 1- μL aliquots in the gas chromatograph, preferably over a period of time rather than in consecutive analyses. If these instructions were followed, 20 ng of each component were injected in the gas chromatograph (GC) for the 50/50 standard, and around 10 and 30 ng were injected for the other standards. If the participant preferred to inject 1.5 μL , then the standard would have been dissolved in 150 μL of solvent.

[7] Sediment samples were obtained from an array of marine locations (Figure 1) to encompass a variety of sea surface temperatures, depositional environments (pelagic to lacustrine), and organic composition (oligotrophic to upwelling). Samples A and E were surface sediments, whereas sample C was from a core section (approximate age of 10,000 years) and samples B and D were obtained from core catchers (sediment age corresponding to isotopic stage 6). Samples A, B, and D were chosen to represent the type of samples that most laboratories typically analyze and to represent deep-sea environments. In contrast, samples C and E were expected to pose an increased analytical challenge owing to their



unusual composition and the low abundance of alkenones for sample C, which a number of laboratories had probably not faced before. Sample C corresponded to a lacustrine phase in the evolution of the Baltic Sea. Its alkenone signature was known to be similar to those found in some modern lacustrine environments. Sample E was collected from the vicinity of the Ebro Delta and was influenced by estuarine conditions. Hence it cannot be considered as representative of pelagic Mediterranean Sea sediments. The origins of the sediments were not revealed to the participants. The only information provided was that each sample was retrieved from a marine environment and from different locations, so that the alkenone distributions and concentrations should be expected to be distinctive. Before the sediments were distributed (~ 6 g sent to each laboratory in a vial) they were finely ground with a pestle and mortar, thoroughly manually homogenized, and placed in glass jars. Each vial was filled directly from the glass jar.

2.2. Statistical Scheme

[8] In the first instance, several graphical representations were performed to explore the type of distributions present as well as the occurrence of errant results (histograms, Youden two-sample diagrams and box plots, in addition to standard x - y plots of the means from each laboratory [Youden and Steiner, 1975; Bauer et al., 1986; Isaacs, 2000; Meier and Zünd, 2000; J. J. Filliben and A. Heckert, Web site <http://www.itl.nist.gov/div898/handbook>]). Outliers were then removed from the data set using a combination of graphical and numerical methods (see Table 1), and only then were the rest of statistical calculations carried out. Finally, the similarity of the data and their repeatability and reproducibility were investigated using one-way analysis of variance (ANOVA) coefficients and post hoc tests (Tukey). This analytical scheme is equivalent to the one commonly used

in interlaboratory analytical studies for both geological or nongeological samples, which follow (to different extents) the *International Standardisation Organisation (ISO)* [1994] standard 5725 and the International Harmonized Protocol endorsed by *IUPAC* [1995] (e.g., Kallischnigg and Müller [1984], Boyer et al. [1985], Bauer et al. [1986]; Colombo [1986], Stephens et al. [1992], Lin [1995], Lopez-Avila et al. [1997], Nilsson et al. [1997], Rossmann et al. [1997], van Bavel et al. [1998], and Isaacs [2000]).

2.3. Calculation of Statistical Parameters

[9] There is some debate as to which is the best approach to estimating the best values from interlaboratory analytical data on geochemical samples [e.g., Colombo, 1986]. In Table 2 the usually considered parameters are shown, that is, weighted average of laboratory means (weights are the reciprocals of variances associated with each laboratory mean), the median of all measurements (minus the outliers; procedure contained in the ISO 5725 norm), the mean, and the mode. The dispersion of the data has been calculated using the corresponding statistical parameters to the median (i.e., interquartile range) and the mean (i.e., standard deviation). In addition, using the analysis of variance technique, the interlaboratory variance (reproducibility, s_R) and the intralaboratory variance (repeatability, s_r) have been estimated [Bauer et al., 1986; ISO, 1994; Nilsson et al., 1997; Wood, 1999]. Values of variance have also been expressed in the form of relative standard deviation (RSD, equivalent to the coefficient of variation, percentage units). The reproducibility is the among-laboratories precision. The repeatability is an estimate of the reliability of a method from a particular laboratory [Nilsson et al., 1997] and reflects the precision from the analysis of replicate test samples. The value obtained of within-laboratory variability can be assumed



Table 1. Summary of the Outliers Tests^a

| Laboratory Codes | Standards–U ₃₇ ^{K'} | | | Sediments–U ₃₇ ^{K'} | | | | | Sediments–Alkenones Concentration | | | | |
|------------------|---|----------------|----------------|---|----------------|--------------------------|----|----------------------|-----------------------------------|----------------|----------------|--------------------|----|
| | 3/1 | 1/1 | 1/3 | A | B | C | D | E | A | B | C | D | E |
| 1 | | | | | | | | C | | | | | |
| 2 | | | C | | | | | | | H | | | |
| 3 | | | | | | | | C | | | | | |
| 4 | (BGHC) | (HC) | C | nd | nd | nd (HC ^b) | nd | nd | nd | nd | nd | nd | nd |
| 5 | | | | | | nd | | | | (BHG) | (BHG) | | |
| 6 | C ² | C | HC | | | C | | nd | nd | nd | nd | nd | nd |
| 7 | | | | | | C | C | | | H | | | |
| 8 | | | | | | (HC ^b) | | | | | | | |
| 9 | H | C | | (BGH) | | | | | nd | nd | nd | nd | nd |
| 10 | | | | | | | | | nd | H | | | |
| 11 | | C | | | | | | | nd | nd | nd | nd | nd |
| 12 | | | | | | | | | | | | | |
| 14 | C ¹ | C | nd | C ¹ | C ¹ | (HC ^{1b}) | nd | nd | C ¹ | C ¹ | C ¹ | nd | nd |
| 16 | | H | | (BGHC) | | C | | | | (HC) | | (HC ²) | C |
| 17 | | C ¹ | C ¹ | | | | C | (BGHC) | | | | | |
| 18 | | | | H | (BGH) | (HC ^b) | H | (BGHC) | | | (BHG) | | |
| 20 | C | | | (BGH) | | nd | | (BGHC ²) | | H | nd | | |
| 21 | | | | | | nd | | | | | nd | | |
| 22 | | | | | | C | | | (HC) | (HC) | | (HC) | C |
| 23 | | C | | (BH) | | nd | | | | | nd | | |
| 24 | | | | | | | | | | | | | |
| 25 | | | | | | (GHC) | | | (BHC) | | (BGHC) | | |
| 27 | C | | | | | nd | | | nd | nd | nd | nd | nd |
| 30 | | C | C | | | nd | | | nd | nd | nd | nd | nd |
| Total | 1 | 1 | 1 | 4 | 1 | 5 | 0 | 3 | 2 | 3 | 3 | 2 | 0 |

^aMethods used: B, box plots; G, Grubbs' test, 95% confidence level, used in the Harmonised Protocol [IUPAC, 1995] and the ISO 5725 norm [Grubbs, 1969]; H, Huber's test, lowest possible rejection factor (3) chosen [Davies, 1988; Meier and Zünd, 2000]; C, χ^2 test, 99.95% confidence level, anomalous values detected on the basis of the square distance of a individual mean from the weighted mean and the value of individual standard deviations. C¹ and C² indicate those labs where the χ^2 was not calculated as only one or two replicates were measured and the outcome of the calculation was distorted by an extreme standard deviation. The nd indicates when no data were available for the analysis. Results that were deemed outliers are those that gave positive values in more than one test; these are indicated by parentheses.

^bResults considered outliers on the basis of the histogram distribution and not included in Grubbs' test.



Table 2. Statistical Parameters of the Results From the Analysis of the Test Samples^a

| Sample | Mean ^b | Mean | Weighted Mean | Median | Mode | SD ^b | SD | IR | Expected Value |
|---|-------------------|-------|---------------|--------|-------|-----------------|-------|-------|----------------|
| <i>U₃₇^{K'} Values</i> | | | | | | | | | |
| Standard 3/1 | 0.237 | 0.235 | 0.247 | 0.232 | 0.231 | 0.014 | 0.011 | 0.014 | 0.228 |
| Standard 1/1 | 0.487 | 0.486 | 0.482 | 0.483 | 0.479 | 0.015 | 0.015 | 0.019 | 0.470 |
| Standard 1/3 | 0.742 | 0.741 | 0.745 | 0.740 | 0.741 | 0.017 | 0.016 | 0.021 | 0.726 |
| Sediment A | 0.514 | 0.526 | 0.516 | 0.525 | 0.510 | 0.063 | 0.015 | 0.018 | |
| Sediment B | 0.649 | 0.652 | 0.667 | 0.653 | 0.667 | 0.021 | 0.017 | 0.024 | |
| Sediment C | 0.371 | 0.241 | 0.188 | 0.239 | 0.206 | 0.218 | 0.055 | 0.044 | |
| Sediment D | 0.816 | 0.816 | 0.822 | 0.826 | 0.824 | 0.021 | 0.021 | 0.029 | |
| Sediment E | 0.297 | 0.228 | 0.205 | 0.225 | 0.235 | 0.178 | 0.022 | 0.021 | |
| <i>Alkenone Concentrations, µg/g</i> | | | | | | | | | |
| Sediment A | 0.10 | 0.10 | 0.11 | 0.10 | 0.12 | 0.04 | 0.03 | 0.04 | |
| Sediment B | 7.40 | 7.47 | 7.58 | 7.86 | 8.5 | 3.68 | 2.02 | 1.72 | |
| Sediment C | 0.27 | 0.13 | 0.14 | 0.15 | 0.18 | 0.33 | 0.04 | 0.04 | |
| Sediment D | 1.16 | 1.30 | 1.35 | 1.30 | 1.32 | 0.50 | 0.36 | 0.44 | |
| Sediment E | 1.29 | 1.29 | 0.79 | 1.21 | 1.75 | 0.59 | 0.59 | 0.67 | |

^aAbbreviations stand for SD, standard deviation; IR, interquartile range.

^bWith outliers.

to be overoptimistic if the determinations have been conducted concurrently [Horwitz, 1994]. The repeatability and reproducibility values have also been expressed in the internationally recognized form (95% confidence intervals). Thus the repeatability limit r is the value below which the absolute difference between two single results obtained under repeatability conditions may be expected to lie with a probability of 95%:

$$r = 2.8s_r.$$

Similarly, the reproducibility limit R is

$$R = 2.8s_R.$$

Repeatability and reproducibility are thus defined as the closeness of agreement between two measurements obtained under repeatability and reproducibility conditions, respectively [Nilsson *et al.*, 1997].

3. Results and Discussion

3.1. Analysis of $U_{37}^{K'}$ in Standards

[10] The aim of this part of the study was to test the accuracy and precision of the laboratories'

gas chromatographic procedures to measure $U_{37}^{K'}$. Thus no sample manipulation was required other than the dissolution of the standard mixtures in an appropriate solvent and injection of aliquots into the gas chromatographic system to quantify the synthetic alkenones. The standards should have also provided a reference against which accuracy could be appraised.

[11] After an initial inspection of the data it is clear that most of the results are higher than the expected value and that the distribution of the data tails toward higher values, particularly for the standard "3/1" (Figures 2 and 3). The median, mode, mean, and weighted mean of all values are higher than the expected value for the three standards (Table 2). It is apparent that the relative discrepancy from the expected value tends to have the same sign (positive, i.e., higher and thus warmer) for each laboratory. An analysis of variance (and post hoc test) was conducted that showed that there are laboratories whose results always differ significantly from the expected value (Table A1). Thus there is a bias in the results toward warmer values, the magnitude of which is

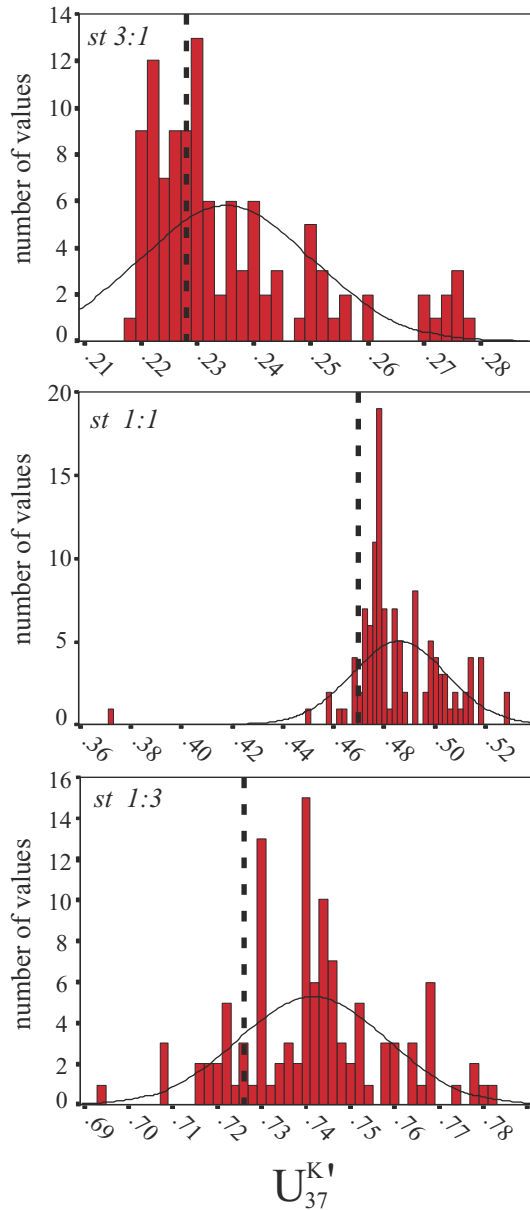


Figure 2. Histograms of all replicate measurements from each laboratory for the analysis of $U_{37}^{K'}$ in the alkenone standards. The interval of $U_{37}^{K'}$ covered by each bar is 0.002. The solid curve represents the normal distribution of the data. The vertical dotted line indicates the expected value of $U_{37}^{K'}$ for each standard.

variable and depends on the laboratory. The magnitude of the bias is such that it may lead to climatically significant errors (e.g., higher than 1°C , although it is clear that smaller values are also climatically important but less so if the glacial/interglacial temperature difference is used as a benchmark; Figure 3).

[12] The cause of such a bias is unclear. The difference observed between the estimates of the best values and the expected value is of the same sign for each standard (Table 2). In fact, there is a perfect linear correlation between the expected value and its difference with the mode ($\Delta(\text{mode-exp value}) \cdot 1000 = -2.4 + 24.1 \text{ exp value}$, $r^2 = 0.999$, $n = 3$). One obvious possibility is that there was an error in the preparation of standards. This would give rise to a general systematic shift in all results, of the same magnitude in the absence of errors specific to a given laboratory, if the error was committed during weighing and/or preparation and mixing of solutions. Impurities in the standards are another possible source of uncertainty. In this instance, however, the bias toward higher values from the expected value would increase more quickly and not linearly as observed (if the mode is indeed providing an accurate estimation of the real value). However, all these hypothetical sources of errors do not account for the different magnitudes of the systematic shift. Thus some laboratories' values are systematically higher than those of other laboratories, for all the standards (notably laboratories 4, 6, 9, 16 and 27). In this case, an analytical factor intrinsic to the chromatographic system of those laboratories may be responsible, which underestimates the $C_{37:3}$ ketone concentration or overestimates that of the other component. Previous studies have argued that the former is more likely to occur when the concentration of the standards is relatively low (dependent on the system). In this case, the $C_{37:3}$ alkenone might be preferentially adsorbed to a component of the gas

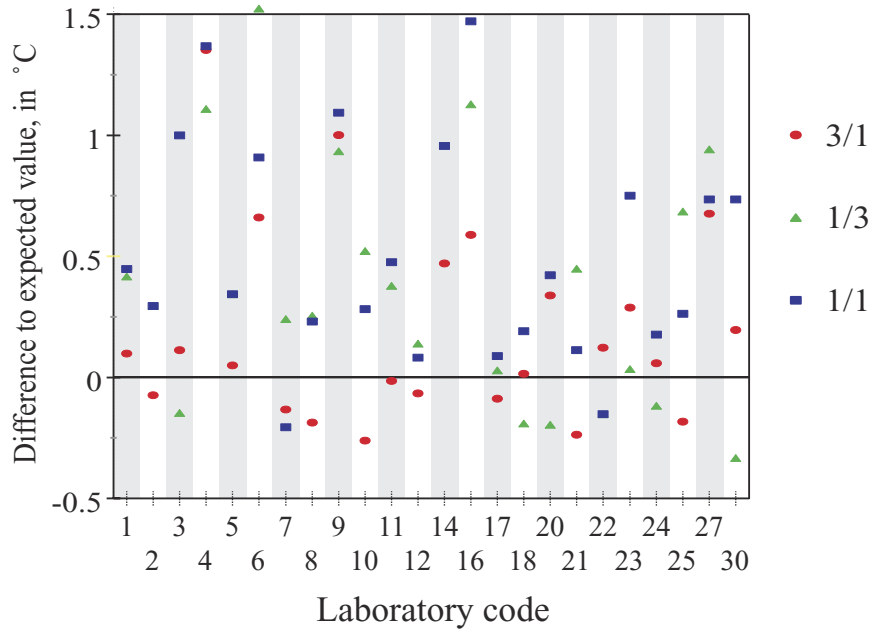


Figure 3. Compilation of the results from the analysis of the three alkenone standards. The value represented in the vertical axis for each laboratory corresponds to the difference between its mean value and the expected value for the standard, converted to degrees Celsius using the calibration equation of $U_{37}^{K'}$ [Müller *et al.*, 1998].

chromatograph [Rosell-Melé, 1994; Rosell-Melé *et al.*, 1995; Villanueva, 1996; Villanueva and Grimalt, 1996].

[13] The repeatability and reproducibility of the results is different for the three standards. The lowest values of s_r and s_R dispersion were obtained for the standard 3/1, and then “1/1” (Table 3). It appears that the measurement of $U_{37}^{K'}$ in the 1/3 standard poses more challenges in terms of precision, and perhaps accuracy. The absolute values of the coefficients are not negligible either. The values of repeatability limit (r) and reproducibility limit (R) can be converted to degrees Celsius to obtain a temperature estimate of the value below which the absolute difference between two single results obtained under repeatability or reproducibility conditions may be expected to lie with a confidence level of 95% (Table 3). These values can be used to appraise the comparability of the

data obtained by different laboratories. The mean value of r (i.e., maximum absolute difference between two single results obtained by the same laboratory, with a confidence level of 95%) converted to ΔT is 0.7°C . The mean reproducibility (R , i.e., maximum absolute difference between results obtained by two laboratories, with a confidence level of 95%) is 1.3°C , higher than the repeatability as might be expected. Certainly, the magnitude of R is large and attests to the relative spread of the mean values of $U_{37}^{K'}$ in Figures 2 and 3. It should be noted, however, that the high value of reproducibility is caused by a few laboratories that systematically produce significantly higher values of $U_{37}^{K'}$.

3.2. Analysis of $U_{37}^{K'}$ in Sediments

[14] As expected, the analysis of the samples posed a varying degree of difficulty, which is

Table 3. Values of Repeatability and Reproducibility Based on the ANOVA Coefficients^a

| | s_r | r | s_R | R | RSD _r , % | RSD _R , % | HORRAT |
|---------------------|---|--------------------------|-------|--------------------------|----------------------|----------------------|--------|
| | <i>U₃₇^{K'} Values</i> | | | | | | |
| Standard 3/1 | 0.005 | 0.014 | 0.012 | 0.035 | 2.2 | 5.3 | |
| Standard 1/1 | 0.007 | 0.021 | 0.016 | 0.045 | 1.5 | 3.3 | |
| Standard 1/3 | 0.012 | 0.033 | 0.018 | 0.051 | 1.6 | 2.5 | |
| Mean converted to T | | 0.7° | | 1.3° | | | |
| Sediment A | 0.028 | 0.079 | 0.029 | 0.082 | 5.4 | 5.6 | |
| Sediment B | 0.018 | 0.049 | 0.023 | 0.065 | 2.7 | 3.6 | |
| Sediment C | 0.027 | 0.076 | 0.061 | 0.169 | 11.3 | 25.1 | |
| Sediment D | 0.013 | 0.036 | 0.025 | 0.069 | 1.6 | 3.0 | |
| Sediment E | 0.017 | 0.047 | 0.025 | 0.070 | 7.3 | 10.9 | |
| Mean converted to T | | 1.7° (1.6°) ^b | | 2.7° (2.1°) ^b | | | |
| | <i>Alkenone Concentrations</i> | | | | | | |
| Sediment A | 0.027 | 0.075 | 0.042 | 0.117 | 26.7 | 41.7 | 5.2 |
| Sediment B | 1.479 | 4.141 | 2.458 | 6.882 | 19.8 | 32.9 | 7.9 |
| Sediment C | 0.026 | 0.074 | 0.035 | 0.097 | 20.2 | 26.7 | 3.5 |
| Sediment D | 0.342 | 0.958 | 0.376 | 1.053 | 26.3 | 28.9 | 5.3 |
| Sediment E | 0.342 | 0.958 | 0.419 | 1.174 | 26.5 | 32.5 | 6.0 |

^aRepeatability variance, s_r ; limit, r ; and relative standard deviation, RSD_r. Reproducibility variance, s_R ; limit, R ; and relative standard deviation, RSD_R. Values of variance and limit are expressed in $\mu\text{g/g}$ of alkenone concentration, and units of $U_{37}^{K'}$. See the text for the description of the calculation of HORRAT values.

^bExcludes sediment C value of R .

reflected in the variability of the results (Figures 4 and 5). The spread of the data for sample C is very large and follows a bimodal distribution, which is also apparent for the results of sample E (see Figure 4). In contrast, the distribution of the data for samples B and D is skewed and tails toward lower values (Figure 4). Arguably, only the spread of data from sample A follows the normal curve, although some apparent outlying values occur at lower $U_{37}^{K'}$ values. Extreme values were then considered to be outliers, and their identification confirmed by the outliers tests (Table 1). The Grubbs' test failed to identify in sample C the grouping of results at high $U_{37}^{K'}$ as outliers, which in the box plots are also included within the range of "valid" results (Figure 5). These were nevertheless identified by other tests as outliers (Table 1) and can be easily picked up in the Youden plots as odd values away from the main cluster of results (Figure 6). For sample B only one of the results was eventually excluded from further calculations, and none of the

results for sample D were excluded (Table 1). For sample A, in contrast, the results of 4 laboratories out of 23 can be considered as outliers; similarly, for sample E with 3 out of 21 results can be considered outliers, as opposed to sample C with 5 outliers out of 16 results (Table 1).

[15] These results raise the obvious questions as to why some of the results are outliers and why, for samples C and E, do the outlying values have similar values? If a typographical error is excluded, one possibility is inhomogeneity of the test samples. This is unlikely to have been an issue in this study as marine sediments are easily ground and homogenized. Moreover, some laboratories' results have been identified as outliers on more than one occasion (laboratories 18 and 20; see Table 1), with their results being similar to those of other outlying values. It seems quite unlikely that these two laboratories were assigned an odd test sample more than once.

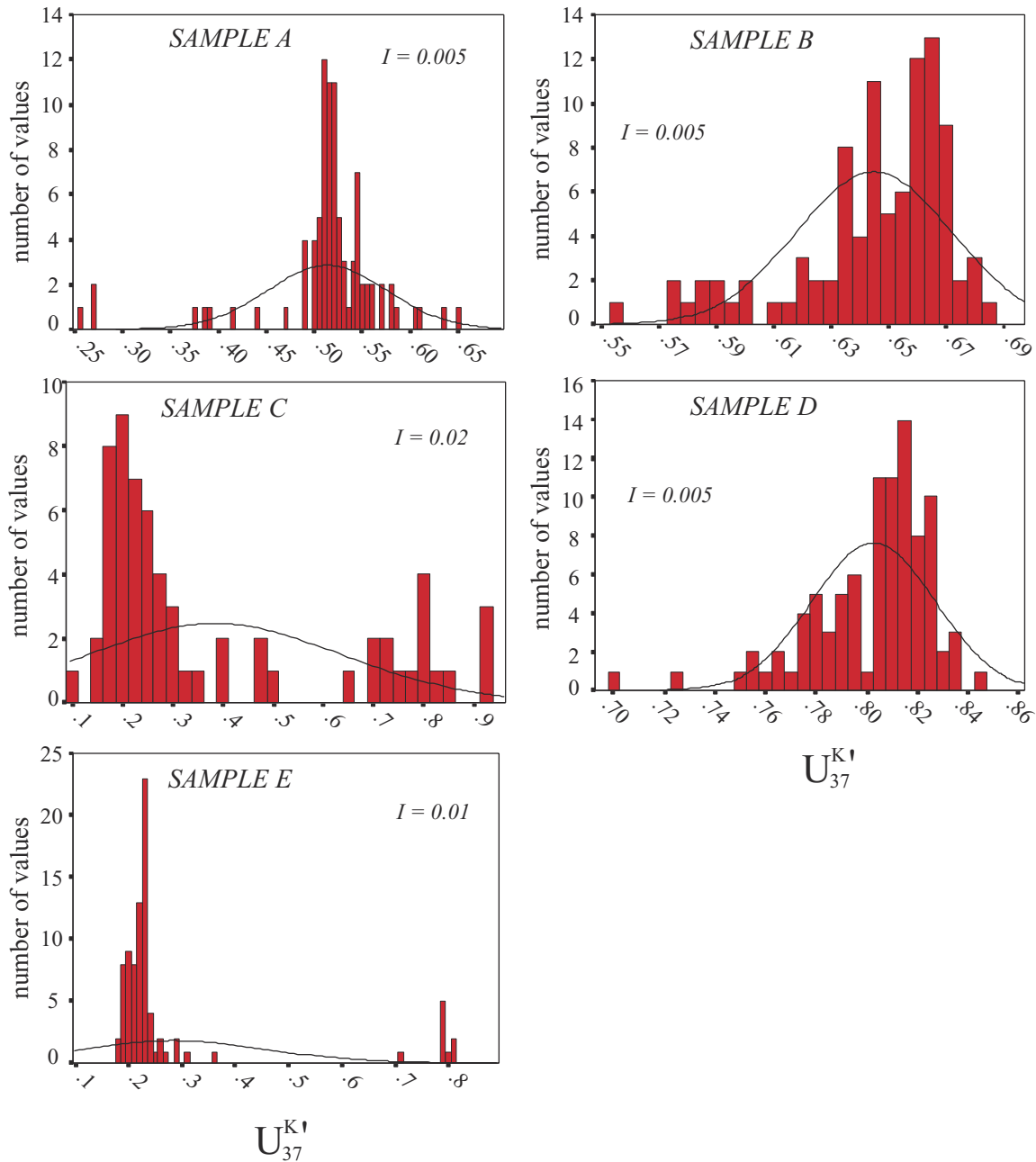


Figure 4. Histograms of all replicate measurements from each laboratory for the analysis of U_{37}^{K1} in the sediment test samples. The interval covered by each bar is indicated in each graph (I). The solid curve represents the normal distribution of the data.

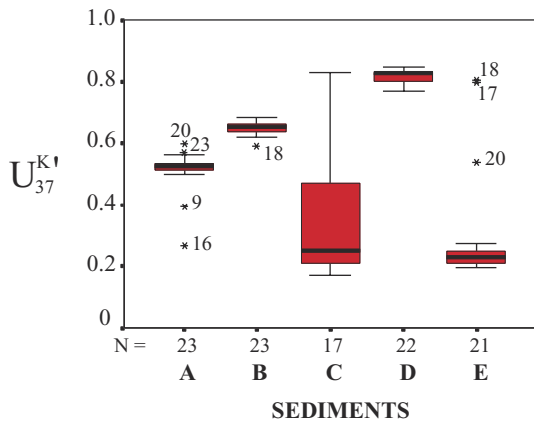


Figure 5. Box plot of the means from each laboratory for the analysis of U_{37}^{K1} in the sediment test samples. The horizontal black line across the box corresponds to the median. Outliers are represented as stars (with the laboratory number) above or below the whiskers. The number of replicate analyses carried out is shown above the sample code.

Hence this suggests that the cause of their outlying result is due to their analytical methodology rather than inhomogeneities of the sediment samples. It is thus possible that the cause of the outlying results in samples C and E, which have similar values, is a common misidentification of the alkenones owing to the unusual distribution of alkenones in both samples. In samples A, B, and D the presence of the $C_{37:4}$ alkenone is very low or below the detection limit, the $C_{37:2}$ component is the most abundant, or as abundant as the $C_{37:3}$ ketone (i.e. sample A), and a similar pattern can also be seen for the C_{38} alkenones which elute a few minutes later (e.g., sample B, Figure 7). In contrast, the relative abundance of alkenones in samples C and E follow a different pattern (Figures 8 and 9), where the concentration of the $C_{37:4}$ alkenone is not negligible, the $C_{37:3}$ alkenone being much more abundant than $C_{37:2}$ (sample E, Figure 9) or as abundant as the $C_{37:4}$ (sample C, Figure 8). Moreover, the C_{38} alkenones for both these samples also present an unusual

distribution pattern owing to the absence of ethyl-substituted homologues and presence of tetra-unsaturated C_{38} ketones. In addition, recognition of the alkenones might have been complicated by the presence of other lipids in the chromatogram. Thus some laboratories might have misidentified the alkenones if they were exclusively reliant on retention times of the chromatographic peaks and visual inspection of the chromatograms.

[16] The ANOVA showed that only for sample A are all the means equal (Table A2). For samples B and D the post hoc analysis (summary of results in Figure 10) reveals that the lowest values of U_{37}^{K1} are significantly different from those in the main cluster of data. In particular, there are three mean values for sample D (laboratories 18, 25, and 30) that are different from more than half of the remaining data (Figure 10). Similarly, for sample C, three mean values are significantly higher than half or more of the results (laboratories 7, 9, and 22). Finally, for sample E, two of the laboratories account for more than 50% of dissimilar results (laboratories 5 and 22). The result of the ANOVA and the post hoc tests is also dependent on the magnitude of the standard deviations of the laboratories' means. Thus, if these are very large, the means are most likely to be comparable for a given difference between two values than if the precision is much higher. It can then be useful to appraise the repeatability and reproducibility of the data to understand the results of the ANOVA and post hoc test (see Table 3). For instance, analysis of sample C led to the largest number of dissimilar results both in terms of the number of outliers and significantly dissimilar means. It is not surprising then that values of interlaboratory variance (s_R) are the highest of the five test samples (Table 4). It is perhaps more surprising that the second highest interlaboratory variance is found for

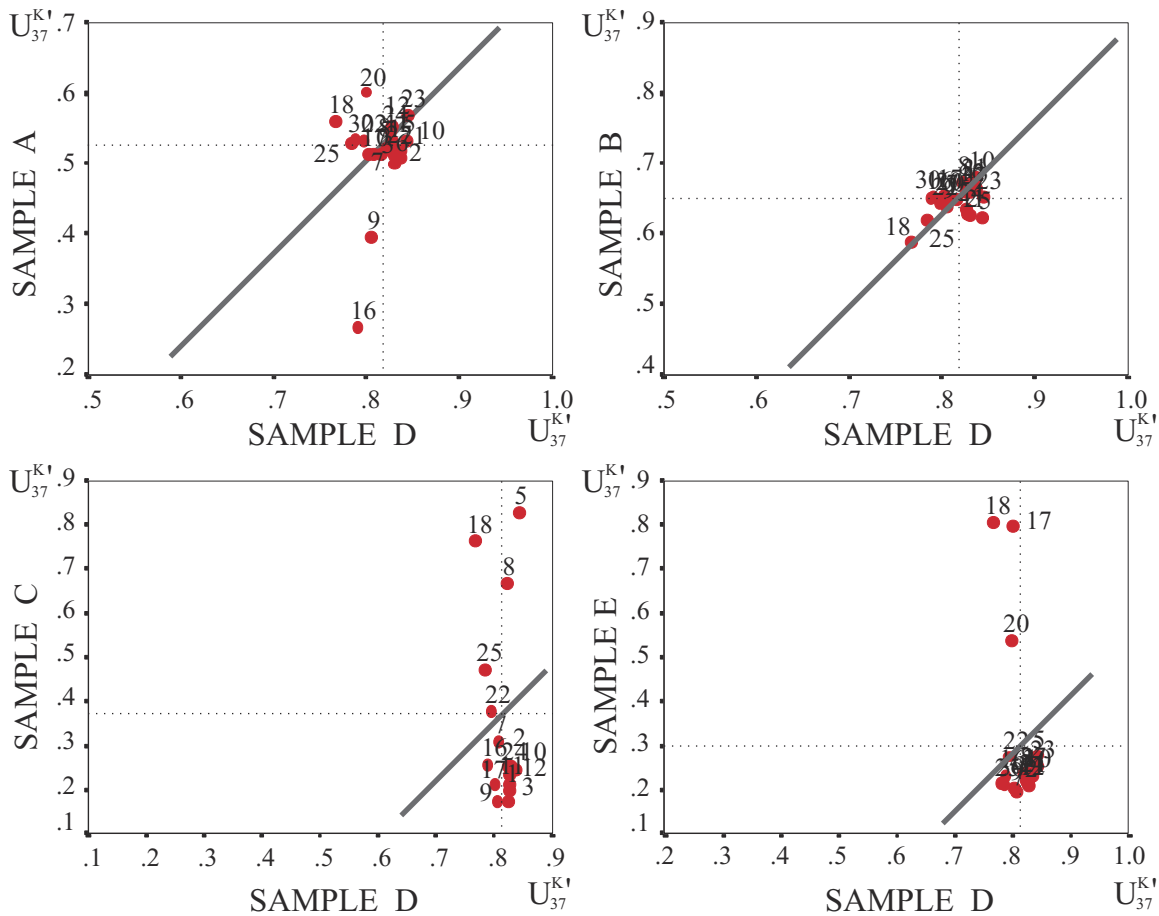


Figure 6. Youden plots of the means from each laboratory for the analysis of U_{37}^{K1} in the sediment test samples. Numbers refer to the laboratories' code. The dotted lines indicate the mean (with outliers) for each sample. The solid diagonal line (45° angle) across the plot is used for reference.

sample A, despite all means being equal as determined by the ANOVA. The latter can be explained by the value of intralaboratory variance (s_r) for sample A, which is the highest of all samples (Table 4). Conversely, the two lowest values of s_r are for samples D and E, both with a higher number of dissimilar results than sample B, which, in turn, has a higher s_r than for both samples D and E. Note as well that the interlaboratory variance for samples B, D, and E is similar (Table 3). Hence it could be argued that if the precision of the results for sample A had been as high

as that for samples B, D, and E, a similar ANOVA result would have been obtained (i.e., not all means significantly similar).

[17] The values of reproducibility are similar for all samples except sample C, which are much higher, i.e., lower reproducibility (Table 3). Certainly, this also illustrates the unusual nature of sample C and that it is probably not comparable to the other samples (i.e., laboratories' methods are not well designed to analyze the type of sample represented by sample C, so that the results from its analysis

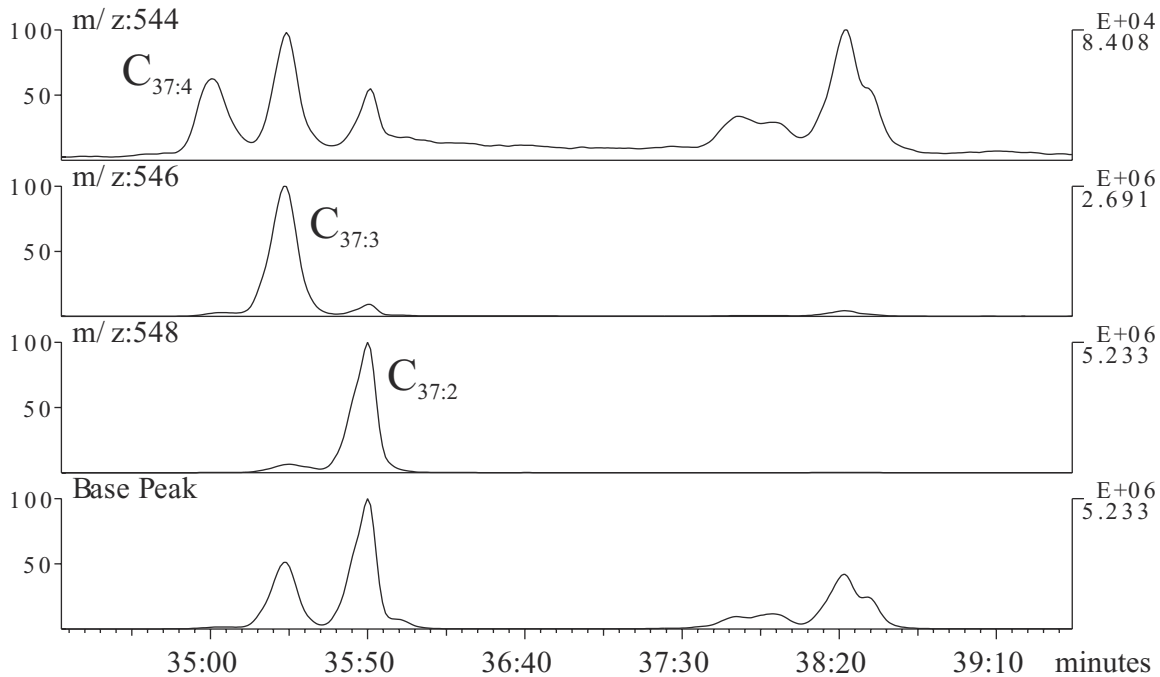


Figure 7. For sample B, ammonia chemical ionization mass chromatograms of the pseudomolecular ions $[M + \text{NH}_4^+]$ of the C_{37} alkenones (description of the method is given by Rosell-Melé *et al.* [1995]). The base peak representation is equivalent to a reconstructed chromatogram with the main ions intensity.

cannot be used to draw conclusions as to the overall capability of a particular laboratory to measure $U_{37}^{K'}$). At present, estimates of temperature produced by different laboratories can differ by up to 2.1°C (or 2.7°C if sample C is taken into account). Clearly, some laboratories will deliver more accurate and precise results than others, and after the completion of the study most laboratories performance should have also improved. Consideration of the value of 2.1°C as the error value of SST estimates between laboratories can be useful, however, to appraise the existence of systematic shifts and the intercomparability of data from different workers, and some of the compilations of data published to date. The high value of R also shows that there is some way to go to improve the accuracy of $U_{37}^{K'}$ estimates. In the present study, accuracy can-

not be appraised because the real value of $U_{37}^{K'}$ is not known, but it is clear that at present, the interlaboratory comparability of the data is low, which must have some bearing on accuracy as well. The repeatability of the results could also be improved and lowered to values of r at least below 1°C , rather than 1.6°C as found in this study, to minimize the occurrence of values that can be interpreted as climatically relevant rather than just analytical noise. There could, however, be a link between the intralaboratory variance and the concentration of alkenones in the sample. Thus the concentrations of alkenones (Table 2) for samples A and C are the lowest and are similar, while their values of s_r (Table 4) are the highest of the sample set. According to these data, if the concentration of alkenones is below 300 ng/g the repeat-

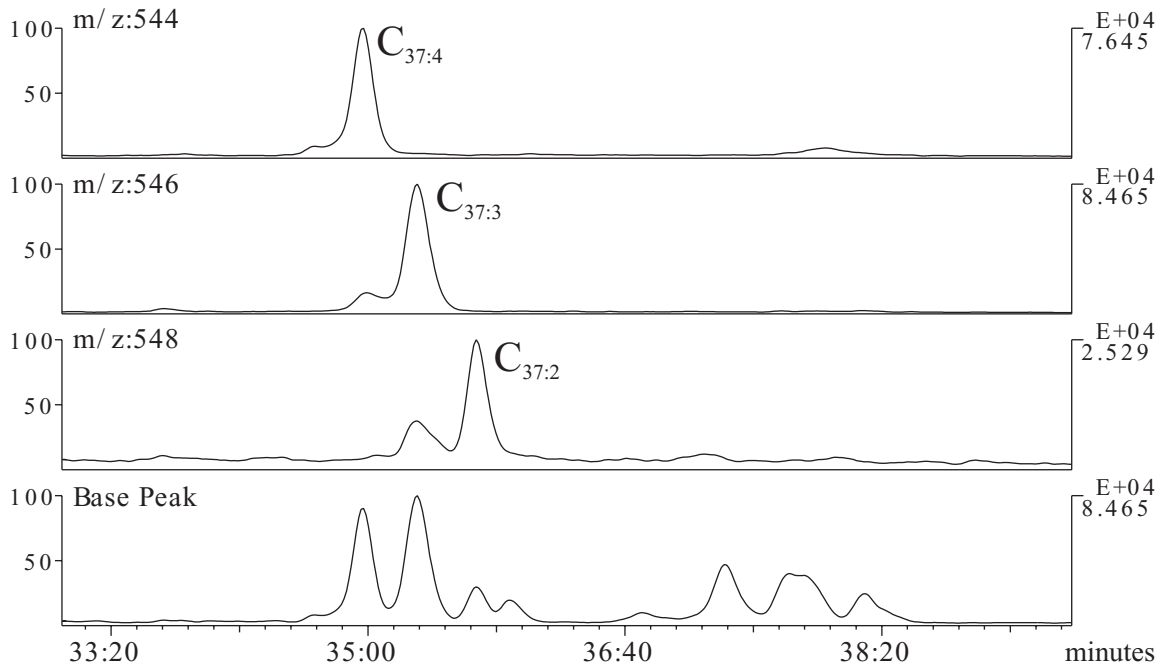


Figure 8. For sample C, ammonia chemical ionization mass chromatograms of the pseudomolecular ions $[M + NH_4^+]$ of the C_{37} alkenones (description of the method is given by Rosell-Melé *et al.* [1995]). The base peak representation is equivalent to a reconstructed chromatogram with the main ions intensity.

ability limit is $>2.2^\circ\text{C}$, whereas if the concentration is higher than $1\ \mu\text{g/g}$, the repeatability is $<1.5^\circ\text{C}$. Hence, in the absence of a detailed assessment of laboratory precision it is recommended that the concentration of alkenones is always provided with published temperature estimates from $U_{37}^{K'}$ to obtain a rough estimate of the precision (i.e., repeatability) of the data.

3.3. Comparison of Methods to Determine $U_{37}^{K'}$

[18] Details of the analytical methodology for some of the laboratories are compiled in Table 4. Note that not all participants volunteered this information. Most studies follow a conventional organic geochemical protocol, which is applicable to the study of a wide range of lipids in sediment samples. The main difference in the protocols lies in per-

forming a sample cleanup prior to chromatographic analysis. Some laboratories used a silica gel column (with variations, e.g., disposable cartridges or glass column with multiple fractions), and in addition, some hydrolyzed the total extract.

[19] To investigate the importance of the cleanup step, data were classified into three groups according to the type of procedure employed (i.e., no cleanup, one or two cleanup steps). The ANOVA of the means of the three procedures (note that a hypothesis test was used for sample C as only one laboratory carried out no cleanup) indicates that only for samples B and E are the means not equivalent. In both cases the use of one cleanup step, with silica gel, provides the discrepant result. Thus arguably the use of only one cleanup step seems to be a more likely source of uncertain results. In fact, the

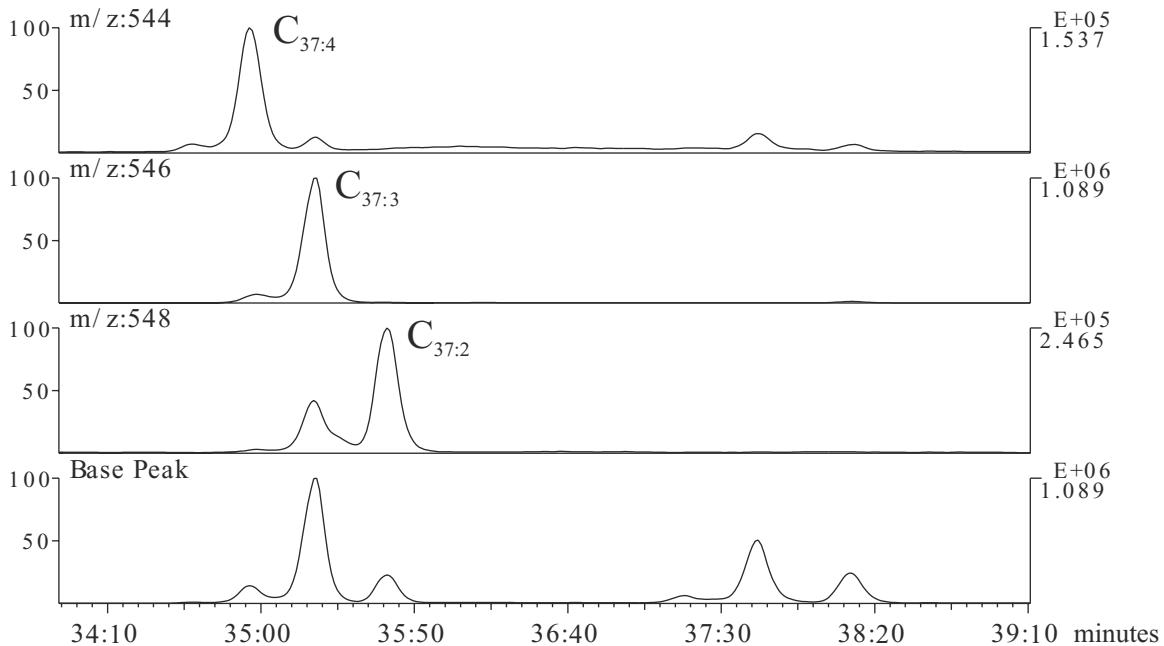


Figure 9. For sample E, ammonia chemical ionization mass chromatograms of the pseudomolecular ions $[M + \text{NH}_4^+]$ of the C_{37} alkenones (description of the method is given by Rosell-Melé *et al.* [1995]). The base peak representation is equivalent to a reconstructed chromatogram with the main ions intensity.

four laboratories that did not use any cleanup step tended to provide a lower number of significantly different results. Hence the use of a cleanup step, per se, does not seem to be essential to derive reliable $U_{37}^{K'}$ values. This, however, may not necessarily mean that such a procedure is unreliable. For instance, not all laboratories using this procedure provided spurious data. Certainly, there are many more factors that have not been appraised that may have a bearing on the quality of the data, such as the skill and experience of the analyst. This may be especially important in the case of the laboratories that appear to inject total aliquots of total extracts directly into the GC. Clearly, after some time the chromatographic performance will degrade as nonvolatile and/or polar components clog up the system, unless action is taken and attention is paid when processing the chro-

matographic data and in maintaining the chromatographic system in good working order.

3.4. Determination of the Concentration of Alkenones in Sediments

[20] The concentration of alkenones in the samples varies by almost 2 orders of magnitude (Table 2). It is apparent that the differences between the results are large (Figures 11 and 12), and hence the ANOVA has showed that not all means are equivalent (Table A2). As in the case for the $U_{37}^{K'}$ results, a few laboratories account for most of the insignificantly similar data (Figure 10). The occurrence of systematically low and high (from the median value) results is prevalent (Figures 12 and 13). This is likely to be a reflection of the different procedures used to quantify the alkenones.

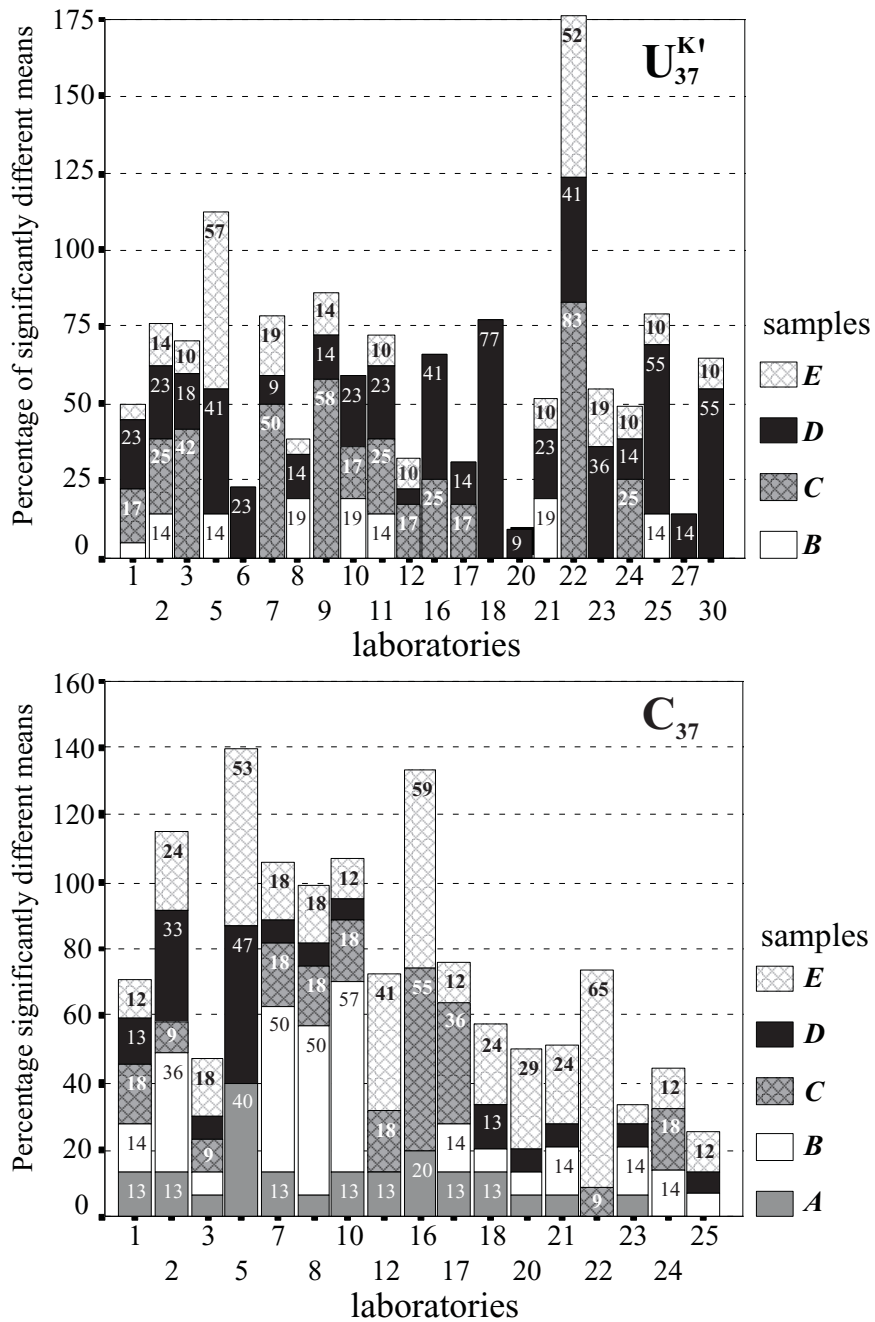


Figure 10. Summary of the Tukey post hoc test (95% confidence) for the analysis of (top) $U_{37}^{K'}$ and (bottom) the concentration of alkenones in the sediment test samples. The x axis refers to the laboratories' code. For each laboratory the height of each colored bar and the number within the bar indicate the percentage of laboratories that have a significantly different mean to the mean of that laboratory. Outliers are excluded from the calculations.



Table 4. Summary of Methods to Measure $U_{37}^{K'}$ for Those Laboratories That Provided the Information^a

| Extraction | Cleanup | | | | Gas Chromatography | | | |
|------------------|---------------|---------------|------------|----------------|--------------------|--------------|---------------------------------|----------|
| | Wash Extracts | Silica Column | Hydrolysis | Derivatization | Injector | Guard Column | Column | Detector |
| Soxhlet | | | | | | Y | DB-1, 60 m, 0.32 × 0.1 | FID |
| Metabolic shaker | | | | | MDGC splitless | | DB-1, 60 m, 0.32 × 0.1 | FID |
| Ultrasound | | | | | on-column | | DB-5, 60 m, 0.25 × 0.1 | FID |
| Ultrasound | Y | Y | | | splitless | | CPSil 5CB, 50 m, 0.32 × 0.1 | FID |
| Soxhlet | | Y 4 samples | | | on-column | | DB-1, 60 m, 0.25 × 0.25 | FID |
| Ultrasound | | Y | | | on-column | | DB5HT, 30 m, 0.32 × | FID |
| Ultrasound | Y | Y | | | on-column | | DB-1, 30 m, 0.25 × 0.25 | FID |
| Ultrasound | | Y | | | on-column | Y | HP5, 50 m, 0.32 × 0.17 | FID |
| Ultrasound | | Y | | Y | on-column/SPI | | HP-1, 60 m, 0.32 × 0.25 | MS |
| Ultrasound | | Y | | | | | DB-5, 30 m, 0.25 × 0.15 | FID |
| Soxhlet | | Y | Y | | Ross | | 25 m, 0.32 × | FID |
| Ultrasound | Y | Y | Y | | on-column | | CPSil 5CB, 30 m, 0.32 × 0.25 | FID |
| Ultrasound | | Y 1 sample | Y | Y | SPI | | CPSil 5CB, 50 m, 0.32 × 0.25 | FID |

^aMDGC, multidimensional chromatography; SPI, septum equipped programmable injector; FID, flame ionisation detector; MS, mass spectrometry; Y, a procedure has been employed by a laboratory.

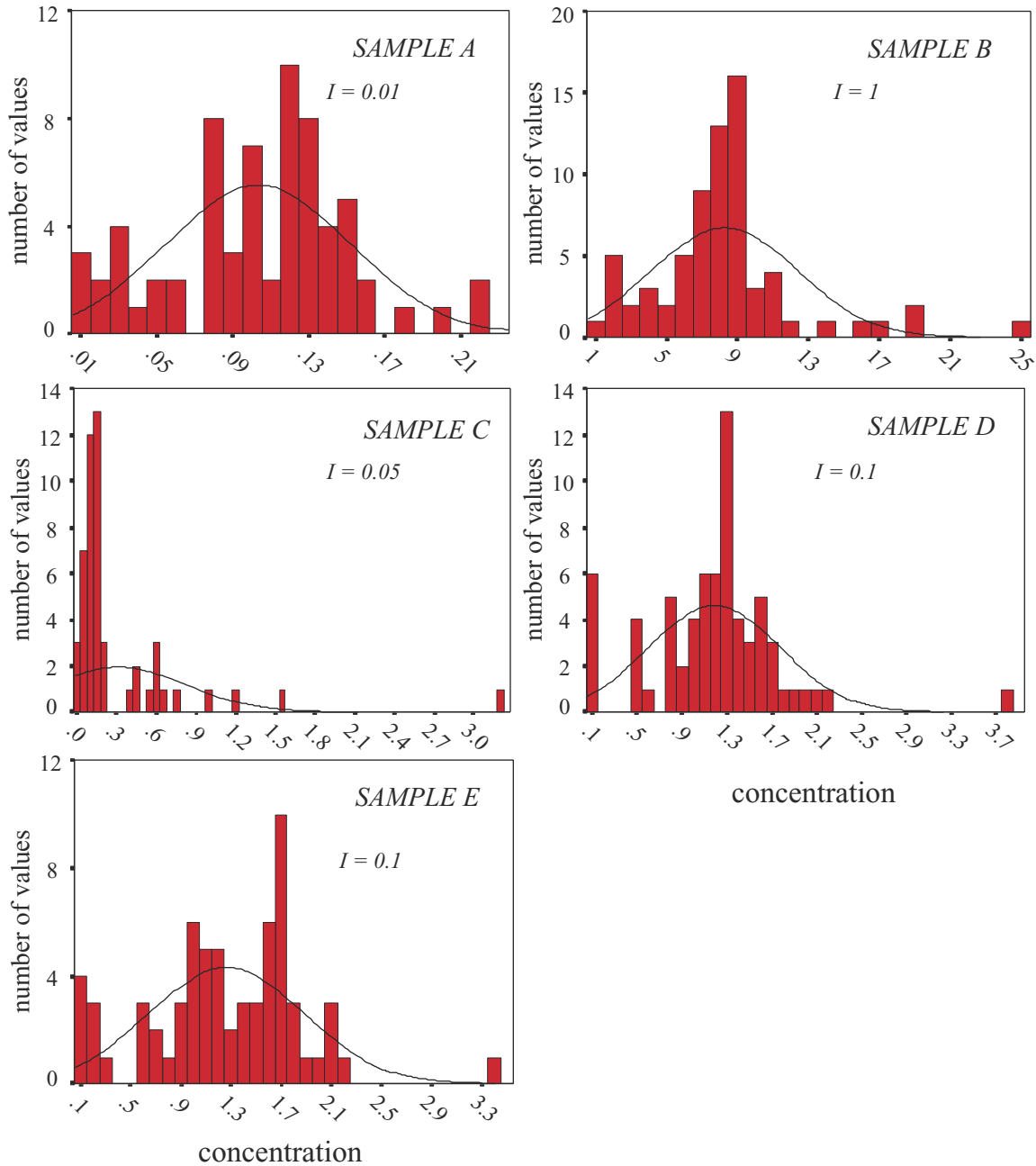


Figure 11. Histograms of all replicate measurements from each laboratory for the analysis of the concentration of alkenones in the sediment test samples. The interval covered by each bar is indicated in each graph (I). The solid curve represents the normal distribution of the data.

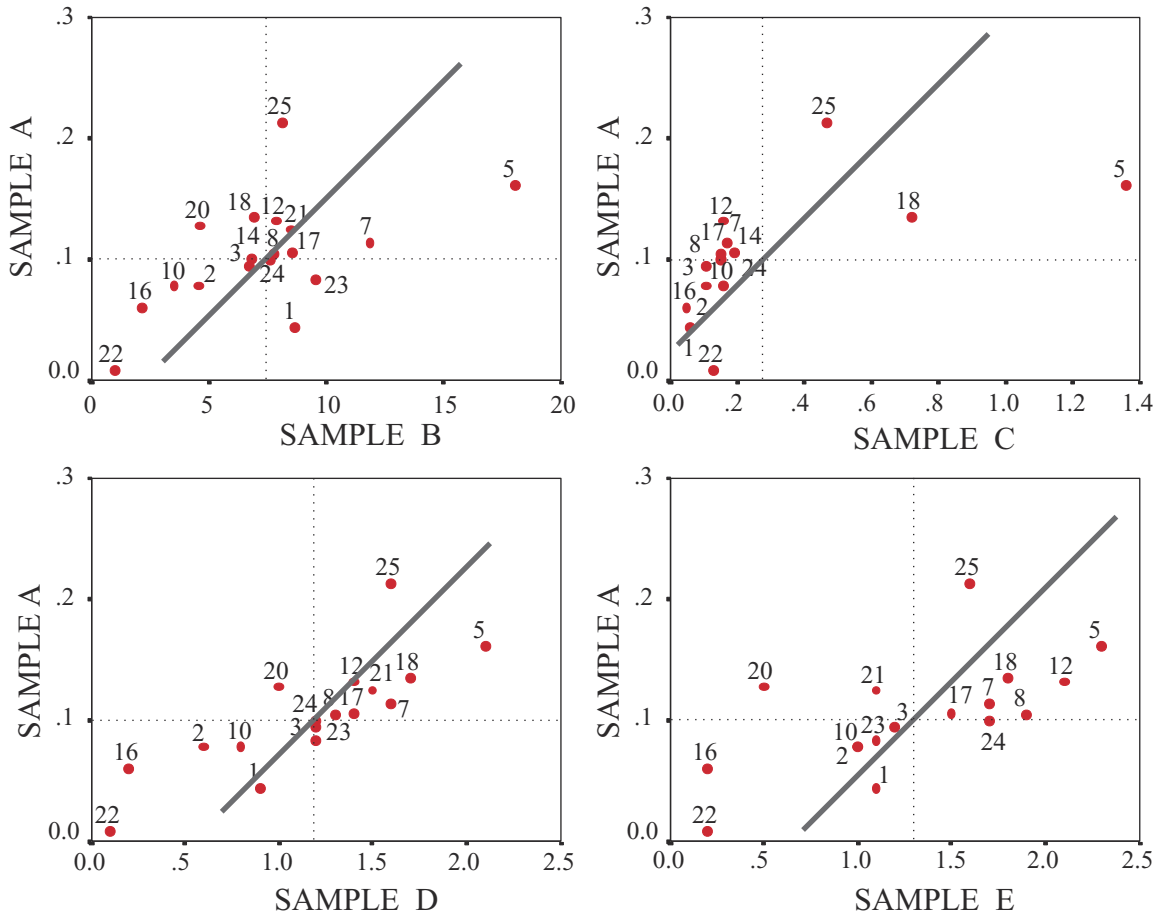


Figure 12. Youden plots of the means from each laboratory for the analysis of the concentration of alkenones in the sediment test samples. Numbers refer to the laboratory code. The dotted lines indicate the mean (with outliers) for each sample. The solid diagonal line (45° angle) across the plot is used for reference.

[21] The dispersion of the data, or the magnitude of the repeatability and reproducibility limits (r and R , respectively), is a function of the concentration of alkenones in the sample (Table 3). Thus it is useful to calculate the relative standard deviation for each dispersion coefficient to compare the results from the samples (see RSD_r and RSD_R , Table 3). It is apparent that the values of RSD_r are comparable for the five sediments and that the slight differences in the values are not related to their concentration. At first, it seems reasonable that

the values of RSD_R are larger than those for RSD_r , and that the former do not appear to be related to concentration. This is, in fact, quite unexpected. The theoretical reproducibility for the methods for measuring alkenone concentrations can be derived from the Horwitz equation [Horwitz, 1982], which reflects the observation that for each hundredfold decrease in analyte concentration an approximately two-fold increase occurs in the RSD. This empirical relationship has been obtained after examining the results of ~3000 collaborative trials and

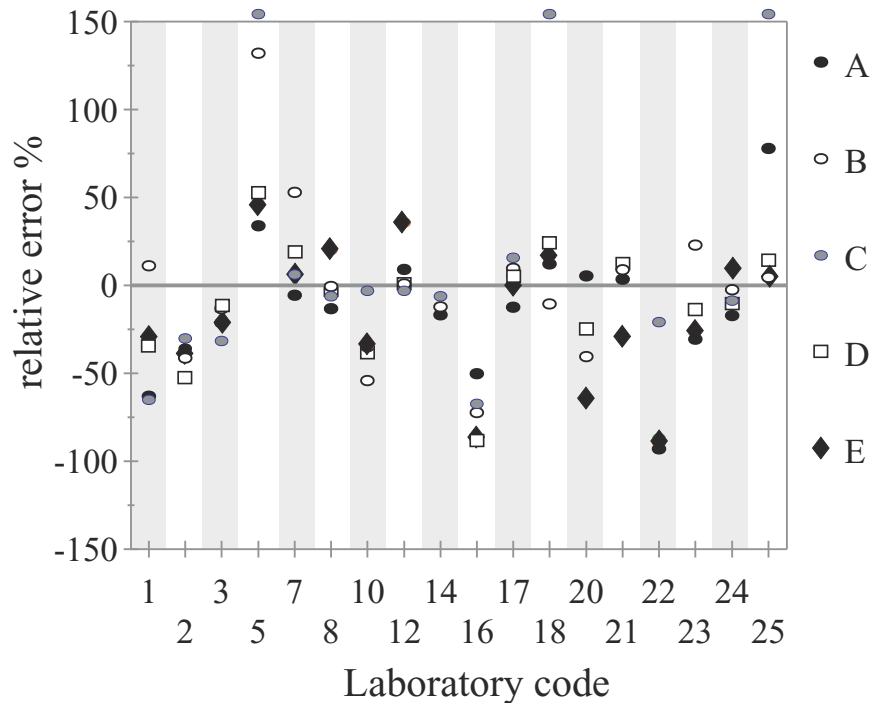


Figure 13. Summary of the results of the analysis of all sediments (A, B, C, D, and E) to determine the absolute concentration of alkenones. The vertical axis represents the difference between the mean value of a laboratory and the median for a particular sample, expressed as a percentage.

establishes that the among-laboratories precision is a function of concentration only, and is independent of analyte, method, and matrix [Boyer *et al.*, 1985]. In this study, HORRAT values (calculated RSD for reproducibility divided by the RSD predicted from the Horwitz equation) [Wood, 1999] have been used to obtain a measure of the acceptability of the among-laboratories dispersion. If the HORRAT values are 2 or less, the method may be assumed to have satisfactory reproducibility values [Boyer *et al.*, 1985; Wood, 1999]. Note, however, that for nonstandardized methods, as in the case of this study, the RSD can be expected to be larger than that predicted by the Horwitz equation [Boyer *et al.*, 1985]. As it turns out, the HORRAT values in this study are much larger than 2 (Table 3). In fact, the analysis of the more concentrated samples

leads to higher values, i.e., much lower reproducibility than expected. So it appears that there is relatively better interlaboratory agreement in the results (i.e., higher reproducibility) from the analysis of samples with a low concentration of alkenones. Of course, this contradicts the observations of Horwitz based in thousands of collaborative trials, and it could be concluded that the reproducibility of the results in the present study is very poor. For instance, for sample B, with the highest concentration of alkenones, the error of calculating the best value is $7.9 \pm 6.9 \mu\text{g/g}$ (based on the reproducibility limit with 95% confidence level). Similarly, for the sample containing the least alkenones (sediment A) the error of the best values is $0.10 \pm 0.12 \mu\text{g/g}$. Thus there is ample scope for improvement in the reduction of the interlaboratory variability. One of



the first objectives should be to reduce the systematic biases between laboratories (related to accuracy). Second, each laboratory should also aim at improving repeatability, which at present is also very low (RSD is $\sim 20\%$). However, in any future attempts to reduce the interlaboratory dispersion in the measurements it may be useful to consider a threshold value below which any discrepancy between two values (i.e., reproducibility limit R) may not be climatically/environmentally significant. Thus, in a sediment core, variations in the sedimentary concentration of alkenones between glacial and interglacial episodes can sometimes vary from several times to orders of magnitudes [e.g., *Madureira et al.*, 1997; *Müller et al.*, 1997; *Rostek et al.*, 1997; *Villanueva et al.*, 1997b; *Schubert et al.*, 1998; *Weaver et al.*, 1999]. Hence it could be argued that the range of errors of the results of the study may be negligible in some instances and that almost all laboratories provided equivalent data. In other circumstances the accuracy of the results may be more critical (e.g., study of Holocene records), so that a consideration of factors that lead to discrepancies in the results is still worthwhile.

4. Conclusions

[22] An important aspect of measuring $U_{37}^{K'}$ and the concentration of alkenones in sediments is to be able to compare reliably data obtained by different laboratories from a wide variety of locations. In this study, such inter-comparability of the data has been achieved by a large number of laboratories within the considered confidence levels, either for the measurement of $U_{37}^{K'}$, or the absolute concentration of alkenones. A number of laboratories need, however, to appraise their procedures, particularly to quantify alkenone concentrations as there are systematic biases between data sets.

[23] Most studies follow a similar analytical protocol for measuring $U_{37}^{K'}$, with the main difference being the use of sample cleanup prior to chromatographic analysis. On the basis of the data available for the methods of the participants in this study that volunteered the relevant information, no preferred method of analysis can be recommended. The use of a cleanup step, however, does not seem to be essential to deriving reliable $U_{37}^{K'}$ values.

[24] From this study it has been estimated that the differences between $U_{37}^{K'}$ temperature estimates from the analysis of oceanic sediment samples, between any two laboratories, may be up to 2.1°C (i.e., 95% confidence level) owing to analytical uncertainties. In addition, the precision of the $U_{37}^{K'}$ temperature estimates from a laboratory is estimated to be as high as $\pm 1.6^{\circ}\text{C}$ (95% confidence level). However, this repeatability limit could be higher ($>2^{\circ}\text{C}$) when the concentration of alkenones in the samples is <300 ng/g (dry weight). Note, however, that the value of this limit does not imply that all data will be randomly distributed around the mean within the limit range. Thus stratigraphic variations in SST of $<2^{\circ}\text{C}$ can be reliably measured, provided, for instance, that the sampling resolution is high enough or if the individual laboratories have lower repeatability limits. In the absence of a detailed assessment of laboratory repeatability it is recommended that the concentration of alkenones is provided with published temperature estimates from $U_{37}^{K'}$ to infer roughly the precision of the measurement. From analysis of alkenone concentrations the interlaboratory reproducibility is estimated to be $\sim 32\%$ and the repeatability is estimated to be $\sim 24\%$. There is certainly scope, and the need, to reduce these values to improve confidence in the data, not just from a particular laboratory but also from the community of scientists involved in alkenone research.



Table A1. Appraisal of the Significance in the Systematic Differences Between the Mean $U_{37}^{K'}$ Obtained by the Laboratories and the Expected Value for Each Standard Using the Tukey Post Hoc Test (95% Confidence)^a

| Laboratory | Standard | | |
|------------|----------|-----|-----|
| | 3/1 | 1/3 | 1/1 |
| 1 | | | |
| 2 | | | |
| 3 | | | T |
| 4 | T | T | T |
| 5 | | | |
| 6 | T | T | |
| 7 | | | |
| 8 | | | |
| 9 | T | T | T |
| 10 | | | |
| 11 | | | |
| 12 | | | |
| 14 | | | |
| 16 | T | T | T |
| 17 | | | |
| 18 | | | |
| 20 | | | |
| 21 | | | |
| 22 | | | |
| 23 | | | |
| 24 | | | |
| 25 | | T | |
| 27 | T | T | |
| 30 | | | |

^aMeans that are significantly different to the expected value are indicated with a T (e.g., all means of laboratory 4).

Table A2. Summary of the ANOVA Results From the Analysis of the Sediment Test Samples^a

| | Sum of Squares | Degrees Freedom | Mean Square | <i>F</i> Measured | <i>F</i> Tabulated | Outcome |
|---|----------------|-----------------|-------------|-------------------|--------------------|-----------------|
| <i>U₃₇^{K'} Values</i> | | | | | | |
| Sediment A | | | | | | |
| Between groups | 1.816E-02 | 17 | 1.068E-03 | 1.349 | 1.92 | equal means |
| Within groups | 4.594E-02 | 58 | 7.921E-04 | | | |
| Sediment B | | | | | | |
| Between groups | 2.567E-02 | 20 | 1.284E-03 | 4.140 | 1.75 | different means |
| Within groups | 2.077E-02 | 67 | 3.100E-04 | | | |
| Sediment C | | | | | | |
| Between groups | 0.125 | 11 | 1.132E-02 | 15.304 | 2.16 | different means |
| Within groups | 2.367E-02 | 32 | 7.397E-04 | | | |
| Sediment D | | | | | | |
| Between groups | 4.171E-02 | 21 | 1.986E-03 | 11.726 | 1.75 | different means |
| Within groups | 1.203E-02 | 71 | 1.694E-04 | | | |
| Sediment E | | | | | | |
| Between groups | 2.826E-02 | 17 | 1.662E-03 | 5.947 | 1.92 | different means |
| Within groups | 1.565E-02 | 56 | 2.795E-04 | | | |

Table A2. (continued)

| | Sum of Squares | Degrees Freedom | Mean Square | <i>F</i> Measured | <i>F</i> Tabulated | Outcome |
|-------------------------------|----------------|-----------------|-------------|-------------------|--------------------|-----------------|
| <i>Alkenone Concentration</i> | | | | | | |
| Sediment A | | | | | | |
| Between groups | 6.776E-02 | 14 | 4.840E-03 | 6.781 | 2.00 | different means |
| Within groups | 3.283E-02 | 46 | 7.137E-04 | | | |
| Sediment B | | | | | | |
| Between groups | 234.520 | 13 | 18.040 | 8.251 | 2.00 | different means |
| Within groups | 96.207 | 44 | 2.187 | | | |
| Sediment C | | | | | | |
| Between groups | 6.946E-02 | 10 | 6.946E-03 | 10.025 | 2.24 | different means |
| Within groups | 1.732E-02 | 25 | 6.929E-04 | | | |
| Sediment D | | | | | | |
| Between groups | 8.758 | 14 | 0.626 | 5.363 | 2.00 | different means |
| Within groups | 5.482 | 47 | 0.117 | | | |
| Sediment E | | | | | | |
| Between groups | 19.169 | 16 | 1.198 | 10.253 | 1.92 | different means |
| Within groups | 5.842 | 50 | 0.117 | | | |

^aRead 1.816E-02 as 1.816×10^{-2} .

Acknowledgments

[25] Ian Harrison, Pau Comes, and Paul Fox are thanked for helping in the organization of this project and for preparing the test samples. Organizational aspects were supported by funds from the EU Environment and Climate Programme under contract ENV4-CT97-0564 to the TEMPUS project.

References

- Ackman, R. G., Fundamental groups in the response of flame ionization detectors to oxygenated aliphatic hydrocarbons, *J. Gas Chromatogr.*, **2**, 173–179, 1964.
- Bauer, C. F., C. L. Grant, and T. F. Jenkins, Interlaboratory evaluation of high-performance liquid chromatographic determination of nitroorganics in munitions plant wastewater, *Anal. Chem.*, **58**, 176–182, 1986.
- Boyer, K. W., W. Horwitz, and R. Albert, Interlaboratory variability in trace element analysis, *Anal. Chem.*, **57**, 454–459, 1985.
- Brassell, S. C., Applications of biomarkers for delineating marine paleoclimatic fluctuations during the Pleistocene, in *Organic Geochemistry Principles and Applications*, edited by M. H. Engel and S. A. Macko, pp. 699–738, Plenum, New York, 1993.
- Brassell, S. C., G. Eglinton, I. T. Marlowe, U. Pflaumann, and M. Sarnthein, Molecular stratigraphy: A new tool for climatic assessment, *Nature*, **320**, 129–133, 1986.
- CLIMAP Project Members, Seasonal reconstructions of the Earth's surface at the Last Glacial Maximum, *Geol. Soc. Am. Map Chart Ser.*, MC-36, 1981.
- Colombo, A., Evaluation of inter-laboratory geochemical data: Some considerations on the Abbey–Rousseau debate, the meaningfulness of the median and other topics, *Geostand. NewsL.*, **10**, 183–189, 1986.
- Comes, P., and A. Rosell-Melé, Modelling of the effect of chromatographic resolution on the determination of the U_{37}^K index, *J. Chromatogr. Sci.*, **37**, 245–250, 1999.
- Dabrio, M. V., F. Farré, M. Gassiot, J. A. García, and R. Martínez, *Cromatografía de Gases I*, 182 pp., Alhambra, Madrid, 1971.
- Davies, P. L., Statistical evaluation of interlaboratory tests, *Fresenius Z. Anal. Chem.*, **331**, 513–519, 1988.
- Grubbs, F. E., Procedures to detect outlying observations in samples, *Technometrics*, **11**, 1–21, 1969.
- Horwitz, W., Evaluation of analytical methods used for regulation of foods and drugs, *Anal. Chem.*, **54**, 67A–76A, 1982.
- Horwitz, W., Nomenclature of interlaboratory analytical studies (IUPAC recommendations 1994), *Pure Appl. Chem.*, **67**, 1904–1911, 1994.
- International Standardisation Organisation, ISO 572, Accuracy (trueness and precision) of measurement methods and results, Geneva, 1994.
- International Union of Pure and Applied Chemistry (IUPAC), Protocol for the design, conduct and interpretation of method performance studies, *Pure Appl. Chem.*, **67**, 331–343, 1995.

- Isaacs, C. M., Statistical evaluation of interlaboratory data from the cooperative Monterey organic geochemistry study, in *The Monterey Formation—From Rocks to Molecules*, edited by C. M. Isaacs and J. Rullkötter, pp. 461–524, Columbia Univ. Press, New York, 2000.
- Kallischnigg, G., and J. Müller, Application of statistical methods in the evaluation of interlaboratory tests, *Fresenius J. Anal. Chem.*, *317*, 241–245, 1984.
- Levitus, S., and T. Boyer, World Ocean Atlas 1994, vol. 4, Temperature, *NOAA Atlas NESDIS 4*, U.S. Dep. of Commer., Washington, D.C., 1994.
- Lin, R., An interlaboratory comparison of vitrinite reflectance measurement, *Org. Geochem.*, *22*, 1–9, 1995.
- Lopez-Avila, V., Y. Liu, and W. F. Beckert, Interlaboratory evaluation of an off-line supercritical fluid extraction and gas chromatography with atomic emission detection method for the determination of organotin compounds in soil and sediments, *J. Chromatogr. A*, *785*, 279–288, 1997.
- Madureira, L. A. S., S. A. van Kreveld, G. Eglinton, M. H. Conte, G. Ganssen, J. E. van Hinte, and J. J. Ottens, Late Quaternary high-resolution biomarker and other sedimentary climate proxies in a northeast Atlantic core, *Paleoceanography*, *12*, 255–269, 1997.
- Meier, P. C., and R. E. Zünd, *Statistical Methods in Analytical Chemistry*, 424 pp., John Wiley, New York, 2000.
- Müller, P. J., M. Cepek, G. Ruhland, and R. R. Schneider, Alkenone and coccolithophorid species changes in Late Quaternary sediments from the Walvis Ridge: Implications for the alkenone paleotemperature method, *Palaeogeogr. Palaeoclimatol. Palaeoecol.*, *135*, 71–96, 1997.
- Müller, P. J., G. Kirst, G. Ruhland, I. von Storch, and A. Rosell-Melé, Calibration of the alkenone paleotemperature index $U_{37}^{K'}$ based on core-tops from the eastern South Atlantic and the global ocean (60°N–60°S), *Geochim. Cosmochim. Acta*, *62*, 1757–1772, 1998.
- Nilsson, T., R. Ferrari, and S. Facchetti, Inter-laboratory studies for the validation of solid-phase microextraction for the quantitative analysis of volatile organic compounds in aqueous samples, *Anal. Chim. Acta*, *356*, 113–123, 1997.
- Prahl, F. G., and L. A. Muehlhausen, Lipid biomarkers as geochemical tools for paleoceanographic study, in *Productivity of the Ocean: Present and Past*, edited by W. H. Berger, V. S. Smetacek, and G. Wefer, pp. 271–289, John Wiley, New York, 1989.
- Prahl, F. G., and S. G. Wakeham, Calibration of unsaturation patterns in long-chain ketone compositions for paleotemperature assessment, *Nature*, *320*, 367–369, 1987.
- Prahl, F. G., L. A. Muehlhausen, and D. I. Zahnle, Further evaluation of long-chain alkenones as indicators of paleoceanographic conditions, *Geochim. Cosmochim. Acta*, *52*, 2303–2310, 1988.
- Rechka, J. A., and J. R. Maxwell, Characterisation of alkenone temperature indicators in sediments and organisms, *Org. Geochem.*, *13*, 727–734, 1988.
- Rosell-Melé, A., Long-chain alkenones, alkyl alkenoates and total pigment abundances as climatic proxy-indicators in the Northeastern Atlantic, Ph.D. dissertation, Univ. of Bristol, Bristol, England, 1994.
- Rosell-Melé, A., J. Carter, A. Parry, and G. Eglinton, Novel procedure for the determination of the $U_{37}^{K'}$ in sediment samples, *Anal. Chem.*, *67*, 1283–1289, 1995.
- Rossmann, A., J. Koziat, G. J. Martin, and M. J. Dennis, Determination of the carbon-13 content of sugars and pulp from fruit juices by isotope-ratio mass spectrometry (internal reference method): A European interlaboratory comparison, *Anal. Chim. Acta*, *340*, 21–29, 1997.
- Rostek, F., E. Bard, L. Beaufort, C. Sonzogni, and G. Ganssen, Sea surface temperature and productivity records for the past 240 kyr in the Arabian Sea, *Deep Sea Res., Part II*, *44*, 1461–1480, 1997.
- Schubert, C. J., J. Villanueva, S. E. Calvert, G. L. Cowie, U. von Rad, H. Schulz, U. Berner, and H. Erlenkeuser, Stable phytoplankton community structure in the Arabian Sea over the past 20,000 years, *Nature*, *394*, 563–566, 1998.
- Sonzogni, C., E. Bard, F. Rostek, A. Lafont, A. Rosell-Melé, and G. Eglinton, Core-top calibration of the alkenone index versus sea surface temperature in the Indian Ocean, *Deep Sea Res., Part II*, *44*, 1445–1460, 1997.
- Stephens, R. D., C. Rappe, D. G. Hayward, M. Nygren, J. Startin, A. Esbøll, J. Carlé, and E. J. Yrjänheikki, World Health Organization international intercalibration study on dioxins and furans in human milk and blood, *Anal. Chem.*, *64*, 3109–3117, 1992.
- Ternois, Y., M. A. Sicre, A. Boireau, M. H. Conte, and G. Eglinton, Evaluation of long-chain alkenones as paleotemperature indicators in the Mediterranean Sea, *Deep Sea Res., Part I*, *44*, 271–286, 1997.
- van Bavel, B., P. Andersson, N. Takeda, and C. Rappe, Results from the second round of the international intercalibration on PCDDs, PCDFs and PCBs in a fly ash extract, *Chemosphere*, *37*, 9–12, 1998.
- Villanueva, J., Estudi de les variacions climàtiques i oceanogràfiques a l'Atlàntic Nord durant els últims 300,000 anys mitjançant l'anàlisi de marcadors moleculars, Ph.D. thesis, Univ. Ramon Llull, Barcelona, Spain, 1996.
- Villanueva, J., and J. O. Grimalt, Pitfalls in the chromatographic determination of the alkenone $U_{37}^{K'}$ index for paleotemperature estimation, *J. Chromatogr. A*, *723*, 285–291, 1996.
- Villanueva, J., C. Pelejero, and J. O. Grimalt, Clean-up procedures for the unbiased estimation of C_{37} alkenone



- sea surface temperatures and terrigenous *n*-alkane inputs in paleoceanography, *J. Chromatogr. A*, *757*, 145–151, 1997a.
- Villanueva, J., J. O. Grimalt, E. Cortijo, L. Vidal, and L. Labeyrie, A biomarker approach to the organic matter deposited in the North Atlantic during the last climatic cycle, *Geochim. Cosmochim. Acta*, *61*, 4633–4646, 1997b.
- Weaver, P. P. E., M. R. Chapman, G. Eglinton, M. Zhao, D. Rutledge, and G. Read, Combined coccolith, foraminiferal, and biomarker reconstruction of paleoceanographic conditions over the last 120 kyr in the northern North Atlantic (59°N, 23°W), *Paleoceanography*, *14*, 336–349, 1999.
- Wood, R., How to validate analytical methods, *Trends Anal. Chem.*, *18*, 624–632, 1999.
- Youden, W. J., and E. H. Steiner, Statistical manual of the association of official analytical chemists: Statistical techniques for collaborative trials, AOAC Int., Gaithersburg, Md., 1975.