



Contents lists available at ScienceDirect

Organic Geochemistry

journal homepage: www.elsevier.com/locate/orggeochem

Performance evaluation of nitrogen isotope ratio determination in marine and lacustrine sediments: An inter-laboratory comparison

Enno Bahlmann^{a,*}, Stefano M. Bernasconi^b, Steven Bouillon^{c,1}, Marco Houtekamer^d, Michael Korntheuer^c, Frauke Langenberg^a, Christoph Mayr^e, Marc Metzke^a, Jack J. Middelburg^d, Birgit Nagel^f, Ulrich Struck^g, Maren Voß^h, Kay-Christian Emeis^a

^aIfBM University of Hamburg, Bundesstr. 55, 20146 Hamburg, Germany

^bGeologisches Institut, ETH Zuerich, Universitätsstr. 16, 8092 Zuerich, Switzerland

^cVrije Universiteit Brussel, Dept. of Analytical and Environmental Chemistry, Pleinlaan 2, B-1050 Brussels, Belgium

^dNetherlands Institute of Ecology (NIOO-KNAW), Korringaweg 7, 4401 NT Yerseke, The Netherlands

^eGeoBio-Center and Department Geo- und Umweltwissenschaften, Ludwig-Maximilians-Universität, Richard-Wagner-Str. 10, 80333 München, Germany

^fInstitute for Coastal Research, GKSS Research Center, Max-Planck-Str. 1, Geesthacht, Germany

^gMuseum für Naturkunde, Humboldt Universität zu Berlin, Invalidenstraße 43, 10115 Berlin, Germany

^hInstitute of Baltic Sea Research, Seestr. 18, 18119 Warnemünde, Germany

ARTICLE INFO

Article history:

Received 28 September 2008

Received in revised form 15 May 2009

Accepted 25 May 2009

Available online xxx

ABSTRACT

Nitrogen isotopes of organic matter are increasingly studied in marine biogeochemistry and geology, plant and animal ecology, and paleoceanography. Here, we present results of an inter-laboratory test on determination of nitrogen isotope ratios in marine and lacustrine sediments. Six different samples covering a wide range of total nitrogen content and $\delta^{15}\text{N}$ values were analyzed by eight different laboratories using their routine procedures. The laboratories were asked to measure three batches with three replicates for each sample to assess accuracy and variability within and among laboratories; this permits assessment of repeatability and reproducibility, which are essential in meta analysis of the increasing database on $\delta^{15}\text{N}$ values in marine sediments. The grand average $\delta^{15}\text{N}$ values for individual samples ranged from 1.65–10.90‰. One laboratory exhibited an average bias of -0.27‰ compared to the mean of all other laboratories. Apart from one sample, which showed an exceptionally high overall standard deviation (OSD) of 0.51‰, the analytical precision (1 s) averaged 0.24‰, ranging from 0.18–0.31‰ for individual samples. Out of the eight participating laboratories, two showed a significantly elevated within-laboratory standard deviation (WLS) of 0.41‰ and 0.32‰ compared to an average WLS of 0.15‰ for the other laboratories. The WLS was inversely correlated with the ratio of peak height to peak width, which was taken as a simple measure of peak shape. Moreover, our data also revealed an inverse correlation between total nitrogen content and measurement precision. These correlations may provide guidance for improving the measurement precision of individual laboratories. Based on the results of this round robin test, we have estimated the expanded measurement uncertainty on the 2σ level to 0.45‰ for sediment samples with a nitrogen content $>0.07\text{ wt}\%$. Sediment samples with lower nitrogen contents cannot be measured with sufficient precision without additional precautions and care should be taken when interpreting $\delta^{15}\text{N}$ signatures and records for sediments with nitrogen concentrations $<0.07\text{ wt}\%$.

© 2009 Elsevier Ltd. All rights reserved.

1. Introduction

The analysis of nitrogen isotope ratios has shown to be valuable in a number of disciplines (e.g., plant and animal ecology, geology, biogeochemistry) and to provide unique insights and essential

information on processes and rates in the global cycle of nitrogen (Fry, 2007). Nitrogen isotopes have been used in studies of organic matter origin (terrestrial versus marine; Peters et al., 1978; Sweeney and Kaplan, 1980) and processing (Wada et al., 1987; Middelburg and Herman, 2007), eutrophication (Owens, 1987; McClelland and Valiela, 1998; Voß et al., 2005), food web studies (DeNiro and Epstein, 1981; Fry, 2007), and in paleoceanography to infer water column denitrification or nitrate utilization from sedimentary records (Altabet, 2007).

This wide applicability of nitrogen isotopes in environmental studies and the widespread availability of automated continuous

* Corresponding author.

E-mail addresses: enno.bahlmann@zmaw.de (E. Bahlmann), kay.emeis@zmaw.de (K.-C. Emeis).

¹ Present address: Katholieke Universiteit Leuven, Dept. of Earth and Environmental Sciences, Kasteelpark Arenberg 20, 3001 Leuven, Belgium.

flow systems in which an elemental analyzer (EA) is coupled directly to an isotope ratio mass spectrometer (IRMS) have resulted in an exponential growth of our database. A prime requisite of using such datasets in meta-analyses, cross-system studies and over large spatial and temporal scales is high accuracy and precision of nitrogen isotope analyses, and confidence in results stemming from different laboratories. In the literature, the precision for $\delta^{15}\text{N}$ determinations is usually given as 0.2‰. Although not (always) clearly stated, this standard deviation (SD) refers to repeatability conditions (within batches of samples, or within laboratories over short time periods). It is often assumed that all IRMS instruments, when calibrated to international standards (i.e., International Atomic Energy Agency (IAEA) standards) and when $\delta^{15}\text{N}$ values are reported relative to atmospheric nitrogen, should in theory yield comparable results for identical well mixed samples. However, the reproducibility (precision among laboratories and over longer periods of time) and thus the comparability of nitrogen isotope analyses is open to debate, and there is mounting evidence that the current reproducibility is not sufficient for meta analysis. [Jardine and Cunjak \(2005\)](#) reviewed the issue of analytical error in ecological studies and identified the need to report overall analytical precision. Moreover, they recommended systematic use of reference materials and conducting blind inter-laboratory comparisons (“round robin tests”).

From $\delta^{13}\text{C}$ round robin tests we know that the reported reproducibility between laboratories was 2–10 times poorer than the sample repeatability ([Verkouteren, 1999](#)). Recently, [Mill et al. \(2008\)](#) reported the within and among laboratory variance of $\delta^{15}\text{N}$ determinations of a single sample (cod muscle) involving eight laboratories. About 74% of the variance could be attributed to differences among laboratories, and laboratory means varied by 2.7‰, revealing high potential bias among laboratories. However, [Mill et al. \(2008\)](#) did not provide essential information on the laboratory procedures and instruments, precluding assessment of potential causes for these poor performances. A round robin inter-laboratory performance test involving various samples might be helpful to identify whether the observations of [Mill et al. \(2008\)](#) are common and whether performance depends on specific laboratory procedures or sample characteristics. For instance, there is increasing evidence, both from literature data as well as from own unpublished data, for poor comparability and poor repeatability of $\delta^{15}\text{N}$ determinations in sediment samples with low nitrogen content. Our current data suggest that carbonate rich matrices may be of special concern, but until now there is a lack of published round robin tests which could substantiate this. Thus, the overall objective of this round robin test is to assess the current analytical status of the $\delta^{15}\text{N}$ determination in natural sediments with a special focus on samples with low nitrogen content.

This inter-laboratory comparison aims:

1. To evaluate the agreement of $\delta^{15}\text{N}$ results in terms of accuracy and within and among laboratory reproducibility,
2. To link any differences (systematic errors or uncertainties) to matrix effects or to the analytical procedure used, and
3. To provide guidance to improve or harmonize our current analytical procedures.

2. Materials and methods

Six different samples were prepared and dispatched to eight laboratories; a number (1–8) was assigned to each laboratory. The laboratories were asked to use their in-house method and calibration procedure and to provide corrected results along with the raw data and a clear description of the method to allow proper

assessment of the data. Corrections, statistical treatment and evaluation of the data have been carried out by IFBM personnel following ISO 5725-2 guidelines for collaborative studies.

2.1. Materials

A suite of six sediment samples was selected which covers a wide range of total nitrogen content and $\delta^{15}\text{N}$ values, but with special emphasis on samples with low nitrogen content (<0.3% dry weight), because isotopic variability is higher at low nitrogen contents ([Hansen and Sommer, 2007](#)).

For each sample, 40–100 g of leftover material from up to 10 different samples with similar characteristics (e.g., material from the top of short sediment cores or surface sediments from the same region) were mixed. Differences in $\delta^{15}\text{N}$ between the leftovers used to produce a mixed sample were generally below 0.4‰. The well mixed samples were freeze dried again and then thoroughly ground with an agate mill. Afterwards, the samples were manually homogenized further. The mixed samples have been analyzed for concentrations of total carbon (TC), total organic carbon (TOC), total nitrogen (TN) and opal, all given in wt% ([Table 1](#)).

Inorganic carbon (IC) was calculated by difference (TC–TOC) and used to estimate CaCO_3 content ($\text{IC} \times 100/12$, assuming negligible dolomite contributions), but for sample B (from a soda lake), IC was used to estimate soda content ($\text{IC} \times 106/12$). Since most of the samples were already ground, no analysis for lithogenic content could be carried out, and the difference between 100% and the sum of all analyzed components was attributed to lithogenic material. Samples were not acidified before analysis, because variable matrix-dependent effects have been reported ([Kennedy et al., 2005](#); [Jaschinski et al., 2008](#)). In a side experiment, for which data were not reported, one laboratory subjected one batch to in situ acidification and ANOVA revealed no significant acidification effect.

2.2. Instrumentation and methods

[Table 2](#) gives an overview of instrumentation and experimental set up for all laboratories participating in the test. All laboratories used a combination of elemental analyzer (EA) coupled to an isotope ratio mass spectrometer (EA–IRMS) for analysis of the samples. The principles of the EA–IRMS are described in detail elsewhere ([Preston and Owens, 1983](#)). Briefly, samples were weighed into a tin or silver capsule (the latter was used only by laboratories 4 and 6) and loaded into a sample carousel for automatic analysis. Samples were dropped into the oxidation reactor maintained at temperatures from 980–1080 °C (see [Table 1](#)), and were combusted in the presence of excess oxygen. Oxidation reactors were filled with chromium(III) oxide and silver plated cobalt oxide to ensure complete oxidation. The combustion products were carried with a helium flow (flow rate ranged from 80–120 ml/min) to a reduction column filled with activated copper (laboratory 6 used a reduction column filled with copper oxide at the top and the bottom, and activated copper in the middle part) that is heated (570–660 °C) for removal of excess oxygen and reduction of NO_x to N_2 . Water vapor was removed from the gas stream either by $\text{Mg}(\text{ClO}_4)_2$ (laboratories 1, 3, 6, 7 and 8) or by P_2O_5 (laboratories 2 and 5). With the exception of laboratories 3 and 6, CO_2 was removed from the gas stream either by a soda lime trap (laboratory 1 and 8) or by a mixed bed trap of ascarite followed by P_2O_5 (laboratory 2 and 5), or Carbosorb followed by $\text{Mg}(\text{ClO}_4)_2$ (laboratory 7). The remaining gas phase products were separated by gas chromatography (for columns and conditions used, see [Table 2](#)) and introduced into the mass spectrometer via an interface. IRMS determines relative abundances of $^{14}\text{N}^{14}\text{N}$ and $^{14}\text{N}^{15}\text{N}$ simultaneously on m/z 28 and m/z 29, respectively, and

Table 1

Characteristics of samples used for the round robin test.

| Sample | Total carbon (%) | Total organic carbon (%) | Total nitrogen (%) | Opal (%) | CaCO ₃ (%) | Lithogenic (%) |
|--------|------------------|--------------------------|--------------------|----------|-----------------------|----------------|
| A | 5.3 | 4.9 | 0.69 | 16.3 | 4.0 | 74.9 |
| B | 3.0 | 1.9 | 0.20 | 21.5 | 10.9 ^a | 65.8 |
| C | 8.3 | 1.6 | 0.13 | 5.7 | 55.6 | 37.1 |
| D | 3.7 | 2.2 | 0.28 | 6.2 | 12.1 | 79.5 |
| E | 8.8 | 0.6 | 0.07 | 5.8 | 68.2 | 25.4 |
| F | 7.7 | 0.2 | 0.02 | 4.9 | 61.9 | 33.0 |

^a Calculated as Na₂CO₃.**Table 2**

Instrumentation and instrumental setup used by participating laboratories.

| Component | Lab 1 | Lab 2 | Lab 3 | Lab 4 | Lab 5 | Lab 6 | Lab 7 | Lab 8 |
|-----------------------|--------------------------------------------|----------------------------------------|------------------------------------|--------------------------------------------|----------------------------------------|------------------------------------|----------------------------------------------|---------------------------------------------------------|
| Autosampler | Thermo Scientific AS128 | LS 200 | Thermo Scientific AS128 | Thermo Scientific AS128 | Thermo MAS 200R | CE Instruments AS 200 | Thermo Scientific MAS 200R | Carlo Erba AS 200 |
| Elemental analyzer | Thermo Scientific Flash EA 1112 | Carlo Erba EA 1108 | Carlo Erba NC2500 | Thermo Scientific Flash EA 1112 | Thermo Flash EA 1112 | CE Instruments NC 2500 | Thermo Scientific Flash EA 1112 | Thermo Scientific Flash EA 1112 |
| Interface | Thermo Scientific ConFlo III | Finnigan ConFlo I | Finnigan MAT ConFlo II | Thermo Scientific ConFlo III | Thermo Scientific ConFlo III | Thermo Finnigan ConFlo II | Thermo Scientific ConFlo 4 | Thermo ConFlo III |
| Mass spectrometer | Thermo Scientific Delta V Advantage | Finnigan Delta S | Finnigan MAT 252 | Thermo Scientific Delta V | Thermo Finnigan Delta Plus XP | Thermo Finnigan Delta Plus | Thermo Scientific MAT 263 | Thermo Delta V Plus |
| Ox-column (°C) | 1010 | 1020 | 1080 | 1020 | 1050 | 1080 | 1020 | 980 |
| Redox-column (°C) | 650 | 650 | 650 | 650 | 570 | 650 | 650 | 640 |
| GC-column | Haysep-Q; 80–100 mesh, 4 m (L) * 4 mm (ID) | Hekatech CN-IRMS HE 26070500 | Hekatech CN-IRMS HE 26070500 | Haysep-Q; 80–100 mesh, 4 m (L) * 4 mm (ID) | Hekatech CN-IRMS HE 26070500 | CE Instruments packed column | Porapak QS 50–80 mesh, (Thermo Fisher) | 4m Porapak Q 50–80 mesh (Elemental Microanalysis E3008) |
| Ash tray | Yes | Yes | No | Yes | No | Not reported | No | Not reported |
| Temperature | 45 | 85 | 60 | 45 | 85 | 45 | 45 | 50 |
| GC-oven (°C) | | | | | | | | |
| Carrier flow (ml/min) | 90 | 80 | 120 | 90 | 120 | 80 | 80 | 110 |
| Water trap | Mg(ClO ₄) ₂ , 10 cm | P ₂ O ₅ | Mg(ClO ₄) ₂ | Mg(ClO ₄) ₂ , 10 cm | P ₂ O ₅ , 10 cm | Mg(ClO ₄) ₂ | Mg(ClO ₄) ₂ | Mg(ClO ₄) ₂ |
| CO ₂ trap | Sodalime | Ascarite/P ₂ O ₅ | – | No trap | Ascarite/P ₂ O ₅ | – | Carbosorb/Mg(ClO ₄) ₂ | Sodalime |

the isotope ratio of nitrogen gas relative to the isotope ratio of a standard or reference gas from a tank with known or predefined isotope ratios.

According to the principle of identical treatment, all participating laboratories performed calibrations on a batch basis either by means of an internationally accepted secondary standard, or by means of a working standard that was checked against an internationally accepted secondary standard on a regular time base. Thus, all reported calibration procedures ensure the traceability of the analytical results to internationally accepted secondary standards, e.g. IAEA N1, IAEA N2, USGS 34 or Schimmelmann acetanilide.

Laboratory 1 performed a two point calibration according to IAEA TECDOC825 page 64, using IAEA N1 or Schimmelmann acetanilide and IAEA N2. Peak areas for *m/z* 28 and *m/z* 29 were integrated manually and corrected for blanks for which a $\delta^{15}\text{N}$ value of 0‰ is stated. For quality assurance two reference standards (sulfanilamide and ISE 964) were measured at the beginning and end of each sequence and monitored by Shewart-charts.

Laboratory 2 used acetanilide as a daily working standard and peptone as quality assurance standard which is regularly analyzed after five samples. The instrument was calibrated with IAEA N1-3 and USGS34 on a regular time base, after maintenance work and after analytical problems. Results were calculated under Isodat 2.0 and corrected against the QS-standard.

Laboratory 3 performed a two point calibration with IAEA N1 and IAEA N2. Each standard has been measured six times within each sequence. The results from Isodat were first corrected for lin-

earity via the peak height and then adjusted against the averages of the calibration standards. The analytical quality was monitored with a sediment quality assurance standard.

Laboratory 4 calibrated with IAEA N1 and IAEA N2 once a month and used a MERCK peptone standard (N-content: 11%; C-content: 44%; $\delta^{15}\text{N} = 7.6\text{‰}$) as a working standard (measured once after five unknowns) to monitor linearity and stability of the instrumentation. No blank correction was applied, but instrument drift was monitored by the peptone standard and was corrected for, if necessary.

Laboratory 5 performed a one point calibration with IAEA N1. During each run four replicates of this standard were measured and the deviation of the mean from the assigned true value was used for correction of the results provided by Isodat. Further, a set of four quality assurance samples (IVA Sediment or IVA Soil) has been measured within each run to monitor the performance of the method and to assess the laboratories long term measurement performance.

Laboratory 6 used peptone as a working standard which was checked against IAEA N1 and IAEA N2 at least every 3–6 months. Along with each batch of 50 samples, 11 peptone standards with different weights and one silver capsule blank at the beginning were analyzed. The elemental standards atropine (C₁₇H₂₃NO₃) and cyclohexanone-2,4-dinitrophenylhydrazone (C₁₂H₁₄N₄O₄) served as further quality assurance standards by laboratory 6. No drift correction or blank correction of the results from Isodat was necessary. For sample F the analytical procedure was slightly modified, in that blanks followed each sample measurement.

Laboratory 7 used atropina, peptone and a sediment standard that were regularly checked against IAEA N1–3 as calibration standards. Calibrations were performed for each batch of samples. No further quality assurance standards were used and no post run correction of the results was performed.

Laboratory 8 calibrated its instrument with IAEA N1 on a daily base. On average 1 standard and 1 blank were analyzed for every 6–10 samples. IAEA N2 was analyzed from time to time to check for consistency with N1. No post run data corrections were applied for the batches reported here.

2.3. Statistical treatment of data

The data have been evaluated following ISO 5725-2 guidelines for collaborative studies. Deviant from this norm, type 1 outliers (intra-laboratory extreme values) were identified based on reported laboratory batch means and not on the base of single analytical results. Because a preliminary analysis of variance (ANOVA) indicated a significant contribution of the between-batch variance to the intra-laboratory variance for individual samples, we found this more suitable.

In the first step of the analysis, extreme values within data from each individual laboratory (type 1 outliers) were identified by means of a Grubb's test. A second Grubb's test was then applied to the remaining data to identify extreme values in the mean of all laboratories for individual samples (type 2 outliers). In both cases, data were tested with a 90% confidence interval. After elimination of type 2 outliers, minimum and maximum batch means of data for all laboratories were calculated for each sample. Type 1 outliers were re-included when lying within these borders, and otherwise eliminated.

Finally, a Cochran's test (99% confidence interval) was applied to exclude outliers related to intra-laboratory variance. In a strict case, the Cochran's test requires equal sample size, but the test is normally rather insensitive to deviations from this premise. Because of the significant contribution of the between-batch variance to the intra-laboratory variance, and because not all laboratories reported complete results, we opted for a two-stage Cochran's test. In the first stage, laboratories that reported data from fewer than two batches for each sample were excluded from the test, and it was only applied to the remaining laboratories. The outliers identified in this first step were excluded. Subsequently, data from labs with incomplete batches were re-included and a Cochran's test was applied again. In addition, the precision of the incomplete datasets was also assessed on batch level. In case the Cochran's test led to an elimination of data, Grubb's test was repeated with the remaining data.

Sample standard deviation (SD) and within-laboratory standard deviations (WLSD) were calculated for both raw data and after elimination of outliers. The bias refers to deviations from the grand averages established by this test after elimination of outliers. For the evaluation of laboratory performance, we calculated bias and WLSD as compared to grand averages. The bias is a measure for the overall trueness of the measurements (systematic errors) and the corresponding standard deviation is used as a measure for the measurement precision.

3. Results

3.1. Database and data quality

A set of six different samples was provided to eight European laboratories for $\delta^{15}\text{N}$ determination. All laboratories were asked to measure each sample nine times (three batches on separate analysis days, with three replicates per batch). Laboratories 1, 2,

3, 5 and 6 have provided complete datasets for all samples. In the dataset provided by laboratory 8, one batch from sample B includes only two measurements. Laboratory 7 has lost one single sample in a batch from sample E and provided only results from one batch for sample F. Laboratory 4 has only reported data from one batch with four replicates for each sample. All laboratories provided a detailed description of their method, calibration procedure and QA measures.

Over all data from all laboratories, 13.5% of the reported results were identified as outliers. With the exception of sample A, all sample datasets contained at least one outlier: Laboratory 4 was excluded from the evaluation of samples B and E due to outliers in the intra-laboratory variance. Moreover, one single value of laboratory 4 was excluded from evaluation of sample D. Laboratory 2 was excluded from the evaluation of sample D, E and F because of extreme values in the laboratory mean, and laboratory 8 was excluded from the evaluation of samples C and E. Batch D-7.2 was excluded from evaluation of sample D as an extreme value within batches. Batches E-6.2, F-6.1 and B-4.1 showed significantly elevated within-batch standard deviations. Because this has no effect on the evaluation results, the data of laboratory 6 were not removed prior to calculation of the grand averages for these samples.

3.2. Sample homogeneity

Six subsamples from each set were randomly selected and approx. 0.5 g of material was taken from these subsamples for homogeneity checks. Subsample homogeneity was assessed by duplicate determination of $\delta^{15}\text{N}$ by laboratory 5. ANOVA indicated no significant contribution of the within sample variance to the total variance for any of the samples; thus, we conclude that the sample material was homogeneous.

3.3. Measurement precision and grand averages

Table 3 provides an overview of the grand average and the assigned measurement precision for all samples. Both the mean and the standard deviation (SD) are given for raw data and after elimination of outliers.

The mean $\delta^{15}\text{N}$ values of the samples ranged from 1.65‰ (sample C) to 10.90‰ (sample B). Elimination of the outliers had no effects on the grand average for samples A, C and D (difference in grand averages <0.02‰), a small effect of +0.05‰ in the case of sample B, and a more distinct effects of >0.10‰ for samples E and F.

The SD for individual samples ranged from 0.18‰ (sample D) to 0.51‰ (sample F) for the raw data; after exclusion of outliers the SD ranged from 0.12‰ (sample D) to 0.38‰ (sample F) (see Table 3). The average SD improved from 0.27‰ to 0.20‰ after removing the outliers. The most prominent effect was observed for samples E and F, where the SD improved by 0.15‰ and 0.13‰, respectively.

3.3.1. Sample A

Sample A is a mix of surface and subsurface multicorer samples taken on AHAB 5 cruise in the upwelling area offshore Namibia between 22°S and 27°S. The sample has the highest TOC and TN values and high opal content (Table 1). The sample data set was free of outliers and revealed a $\delta^{15}\text{N}$ value of 6.74 ± 0.20 ‰. The bias of individual laboratories ranged from -0.15 ‰ (laboratory 1) to $+0.27$ ‰ (laboratory 4), and the within-laboratory standard deviation (WLSD) ranged from 0.07‰ (laboratory 5) to 0.27‰ (laboratory 4).

Table 3
Analytical results and statistics for all participating laboratories.

| | Lab 1 | Lab 2 | Lab 3 | Lab 4 | Lab 5 | Lab 6 | Lab 7 | Lab 8 | All | Without outliers |
|----------|--------------|--------------|--------------|-----------------|--------------|--------------|--------------|--------------|-------|------------------|
| Sample A | | | | | | | | | | |
| Batch 1 | 6.46 ± 0.13 | 6.62 ± 0.14 | 6.90 ± 0.14 | 6.99 ± 0.25 | 6.78 ± 0.05 | 7.04 ± 0.12 | 6.46 ± 0.06 | 6.90 ± 0.03 | | |
| Batch 2 | 6.61 ± 0.1 | 6.61 ± 0.14 | 6.83 ± 0.06 | | 6.77 ± 0.08 | 7.13 ± 0.06 | 6.73 ± 0.06 | 6.55 ± 0.1 | | |
| Batch 3 | 6.63 ± 0.09 | 6.58 ± 0.14 | 6.82 ± 0.10 | | 6.82 ± 0.09 | 6.67 ± 0.10 | 6.59 ± 0.06 | 6.61 ± 0.02 | | |
| Mean | 6.57 | 6.6 | 6.85 | 6.99 | 6.79 | 6.95 | 6.59 | 6.69 | 6.74 | 6.74 |
| n | 9 | 9 | 9 | 4 | 9 | 9 | 9 | 9 | 67 | 67 |
| Lab SD | 0.12 | 0.12 | 0.1 | 0.25 | 0.07 | 0.23 | 0.12 | 0.17 | 0.2 | 0.2 |
| Bias | -0.17 | -0.12 | 0.13 | 0.27 | 0.07 | -0.13 | -0.13 | -0.03 | | |
| Sample B | | | | | | | | | | |
| Batch 1 | 10.84 ± 0.07 | 10.61 ± 0.17 | 11.04 ± 0.06 | 10.25 ± 0.25a,c | 11.03 ± 0.06 | 10.74 ± 0.13 | 10.92 ± 0.07 | 10.84 ± 0.08 | | |
| Batch 2 | 10.83 ± 0.18 | 10.66 ± 0.09 | 11.14 ± 0.04 | | 11.00 ± 0.04 | 10.80 ± 0.04 | 11.06 ± 0.04 | 11.21 ± 0.05 | | |
| Batch 3 | 10.90 ± 0.12 | 10.72 ± 0.12 | 11.00 ± 0.04 | | 11.00 ± 0.03 | 11.00 ± 0.10 | 10.88 ± 0.04 | 10.57 ± 0.08 | | |
| Lab mean | 10.86 | 10.66 | 11.06 | 10.25 ac | 11.01 | 10.84 | 10.95 | 10.83 c | 10.85 | 10.90 |
| Lab SD | 0.12 | 0.12 | 0.07 | 0.25 | 0.05 | 0.14 | 0.09 | 0.27 | 0.24 | 0.16 |
| Bias | -0.03 | -0.22 | 0.17 | -0.64 | 0.12 | 0.06 | 0.06 | -0.06 | | |
| Sample C | | | | | | | | | | |
| Batch 1 | 1.39 ± 0.20 | 1.54 ± 0.13 | 1.71 ± 0.08 | 1.66 ± 0.04 | 1.75 ± 0.01 | 1.55 ± 0.14 | 1.71 ± 0.14 | 2.11 ± 0.11 | | |
| Batch 2 | 1.57 ± 0.08 | 1.21 ± 0.14 | 1.80 ± 0.07 | | 1.72 ± 0.07 | 1.75 ± 0.03 | 1.77 ± 0.12 | 1.54 ± 0.12 | | |
| Batch 3 | 1.67 ± 0.26 | 1.53 ± 0.03 | 1.86 ± 0.17 | | 1.82 ± 0.07 | 1.70 ± 0.22 | 1.66 ± 0.11 | 1.19 ± 0.05 | | |
| Lab mean | 1.54 | 1.43 | 1.79 | 1.66 | 1.76 | 1.67 | 1.71 | 1.61 c | 1.65 | 1.65 |
| n | 9 | 9 | 9 | 4 | 9 | 9 | 9 | 9 | 67 | 58 |
| Lab SD | 0.21 | 0.19 | 0.12 | 0.04 | 0.07 | 0.16 | 0.12 | 0.41 | 0.22 | 0.19 |
| Bias | -0.1 | -0.22 | 0.14 | 0.02 | 0.12 | 0.07 | 0.07 | -0.03 | | |
| Sample D | | | | | | | | | | |
| Batch 1 | 6.63 ± 0.02 | 6.49 ± 0.06 | 6.75 ± 0.07 | 6.93 ± 0.29 | 6.85 ± 0.02 | 6.86 ± 0.04 | 6.90 ± 0.05 | 6.92 ± 0.07 | | |
| Batch 2 | 6.57 ± 0.08 | 6.41 ± 0.01 | 6.84 ± 0.07 | | 6.89 ± 0.09 | 6.91 ± 0.03 | 7.11 ± 0.09a | 6.82 ± 0.09 | | |
| Batch 3 | 6.79 ± 0.05 | 6.55 ± 0.1 | 6.75 ± 0.05 | | 6.96 ± 0.06 | 6.72 ± 0.09 | 6.73 ± 0.01 | 6.74 ± 0.09 | | |
| Lab mean | 6.66 | 6.48 b | 6.78 | 6.93 c | 6.9 | 6.83 | 6.91 a | 6.83 | 6.78 | 6.8 |
| n | 9 | 9 | 9 | 4 | 9 | 9 | 9 | 9 | 67 | 51 |
| Lab SD | 0.11 | 0.09 | 0.07 | 0.29 | 0.07 | 0.1 | 0.17 | 0.11 | 0.18 | 0.12 |
| Bias | -0.14 | -0.32 | -0.02 | 0.13 | 0.1 | 0.11 | 0.11 | 0.03 | | |
| Sample E | | | | | | | | | | |
| Batch 1 | 6.38 ± 0.03 | 5.97 ± 0.09 | 6.26 ± 0.09 | 6.66 ± 0.32 | 6.67 ± 0.05 | 6.63 ± 0.10 | 6.45 ± 0.01 | 5.92 ± 0.06 | | |
| Batch 2 | 6.38 ± 0.03 | 6.05 ± 0.04 | 6.49 ± 0.07 | | 6.65 ± 0.18 | 6.65 ± 0.36 | 6.50 ± 0.07 | 6.46 ± 0.32 | | |
| Batch 3 | 6.70 ± 0.04 | 6.00 v0.11 | 6.44 ± 0.11 | | 6.52 ± 0.06 | 6.47 ± 0.09 | 6.52 ± 0.10 | 5.72 ± 0.15 | | |
| Lab mean | 6.48 | 6.01b | 6.4 | 6.66 c | 6.61 | 6.58 | 6.49 | 6.03 | 6.39 | 6.51 |
| n | 9 | 9 | 9 | 4 | 9 | 9 | 8 | 9 | 66 | 44 |
| Lab SD | 0.16 | 0.08 | 0.13 | 0.32 | 0.12 | 0.21 | 0.07 | 0.38 c | 0.31 | 0.16 |
| Bias | -0.03 | -0.51b | -0.12 | 0.14 | 0.1 | -0.02 | -0.02 | -0.48 | | |
| Sample F | | | | | | | | | | |
| Batch 1 | 4.63 ± 0.14 | 4.30 ± 0.18 | 4.70 ± 0.40 | 4.78 ± 0.14 | 5.26 ± 0.12 | 4.91 ± 0.71 | 4.96 ± 0.04 | 5.74 ± 0.53 | | |
| Batch 2 | 4.35 ± 0.09 | 4.05 ± 0.17 | 5.08 ± 0.09 | | 5.15 ± 0.15 | 4.28 ± 0.30 | n.r. | 5.54 ± 0.18 | | |
| Batch 3 | 4.91 ± 0.10 | 4.34 ± 0.15 | 5.16 ± 0.20 | | 4.44 ± 0.09 | 5.26 ± 0.22 | n.r. | 4.19 ± 0.22 | | |
| Mean | 4.63 | 4.23b | 4.98 | 4.78 | 4.95 | 4.82 | 4.96 | 5.16 | 4.80 | 4.90 |
| n | 9 | 9 | 9 | 4 | 9 | 9 | 3 | 9 | 61 | 52 |
| SD | 0.26 | 0.2 | 0.31 | 0.14 | 0.4 | 0.59 | 0.04 | 0.79 | 0.51 | 0.48 |
| Bias | -0.27 | -0.67 | 0.08 | -0.12 | 0.05 | -0.08 | 0.06 | 0.26 | | |

a: Outlier type 1; b: outlier type 2; c: outlier type 3.

Table 4

Mean deviation from grand average and corresponding standard deviation for individual laboratories participating in the test. The last column denotes the amount (in % of all values) of outliers in the individual laboratory data set as compared to the entire data set.

| Lab ID | Bias (‰) | SD (‰) | n | Outliers (%) |
|--------|----------|--------|-----|--------------|
| 1 | -0.08 | 0.16 | 45 | 0 |
| 2 | -0.27 | 0.16 | 45 | 40 |
| 3 | 0.07 | 0.13 | 45 | 0 |
| 4 | -0.01 | 0.41 | 20 | 45 |
| 5 | 0.11 | 0.09 | 45 | 0 |
| 6 | 0.07 | 0.19 | 45 | 0 |
| 7 | 0.03 | 0.15 | 44 | 0 |
| 8 | -0.11 | 0.32 | 44 | 40.9 |
| All | | 0.23 | 333 | 13.5 |

3.3.2. Sample B

Sample B derives from a tropical soda lake in India, which is strongly influenced by anthropogenic nutrient inputs as reflected

in its high $\delta^{15}\text{N}$ value. This sample is rich in opal; the TOC (1.9%) and TN (0.2%) concentrations are in the mid range covered by the sample set (Table 1).

Individual laboratory means for this sample ranged from 10.25‰ for laboratory 4 to 11.06‰ for laboratory 3 (bias ranged from -0.64–0.12‰); WLSD ranged from 0.04‰ (laboratory 5) to 0.27‰ (laboratory 8). Laboratory 4 and laboratory 8 were excluded from final evaluation of this sample due to extreme values of the WLSD of 0.27‰ and 0.25‰, respectively. The mean of laboratory 4 has also been identified as an outlier within the laboratory means. For this sample, we obtain an uncorrected $\delta^{15}\text{N}$ mean value of $10.85 \pm 0.24\text{‰}$ and a corrected mean $\delta^{15}\text{N}$ value of $10.90 \pm 0.16\text{‰}$.

3.3.3. Sample C

Sample C is an organic rich sediment layer from the Levantine Basin (eastern Mediterranean Sea). The matrix is characterized by relatively high CaCO_3 concentrations (55.6%) and medium range

TOC (1.6%) and TN (0.13%) contents. This sample showed a fairly low $\delta^{15}\text{N}$ value, with a mean $\delta^{15}\text{N}$ value of 1.65‰ for both the uncorrected and the corrected data, and a SD of 0.22‰ including outliers, and 0.19‰ after elimination of data from laboratory 8 (type 3 outlier). The bias of individual laboratories ranged from -0.22‰ (laboratory 2) to $+0.14\text{‰}$ (laboratory 3), and the WLS D ranges from 0.07‰ (laboratory 5) to 0.41‰ (laboratory 8).

3.3.4. Sample D

Sample D consists of pooled sample residues from numerous surface sediment samples from the Skagerrak/northern North Sea. The sample has medium range TOC and TN levels of 2.2% and 0.28%, respectively, and intermediate carbonate and opal contents (Table 1). Batch 2 of laboratory 7 has been identified as a type 1 outlier and the data from laboratory 4 were identified as a type 3 outlier. The uncorrected grand average $\delta^{15}\text{N}$ of this sample was $6.78 \pm 0.18\text{‰}$; individual laboratory means ranged from 6.48‰ (laboratory 2) to 6.91‰ (laboratory 7) and the WLS D ranged from 0.07‰ (laboratories 3 and 5) to 0.17‰ (laboratory 7). Removal of outliers yielded a grand average of 6.80‰ and improved the total SD to 0.12‰.

3.3.5. Sample E

Sample E is a pooled sample, prepared from drill core samples taken at Site 720 and 722 on Ocean Drilling Project Leg 117 on the Oman Margin/Arabian Sea. The sample is rich in carbonate and low in TOC (0.64%) and TN (0.06%). The uncorrected data had a grand average $\delta^{15}\text{N}$ of 6.39‰ and a SD of 0.31‰; laboratory means ranged from 6.01‰ (laboratory 2) to 6.66‰ (laboratory 4), and the WLS D ranged from 0.07‰ (laboratory 7) to 0.38‰ (laboratory 8). The data from laboratory 8 (WLS D of 0.38‰) and laboratory 4 (WLS D of 0.32‰) were identified as outliers in the laboratory variance (type 3 outlier) and were removed from the final evaluation of this sample. After removing these data, the laboratory mean of laboratory 2 became an outlier within the remaining laboratory means and thus was also excluded from the final evaluation which leads to a corrected grand average for sample E of $6.51 \pm 0.16\text{‰}$.

3.3.6. Sample F

Sample F is a multicorer sample from offshore Crete/eastern Mediterranean Sea during r/v Meteor cruise M71/3. This carbonate rich (61.9% CaCO_3) sample has by far the lowest TOC (0.2%) and TN (0.02%) contents of all samples in this study (Table 1). Measurement uncertainty was highest for this sample. The individual bias ranged from -0.62‰ (laboratory 2) to $+0.31\text{‰}$ (laboratory 8) and the WLS D (of the laboratories that provided data from three different batches) ranged from 0.20‰ (laboratory 2) to 0.79‰ (laboratory 8). The uncorrected data yielded a grand average $\delta^{15}\text{N}$ of 4.80‰ with a SD of 0.51‰. After eliminating outliers (laboratory 2: type 2 outlier; laboratory 8: type 3 outlier), the grand average increased by 0.05‰ to 4.85‰ and the SD improved to 0.38‰.

3.4. Factors affecting the measurement precision

3.4.1. Calibration routines

Other authors (Paul et al., 2007) have pointed out that the use of different calibration routines may substantially contribute to the between-laboratory uncertainty and suggested the use of a two or more point calibration routine to minimize this uncertainty. Although there is no doubt that a multipoint calibration is more reliable than a one point calibration, we observed no effect of the different calibration routines used in this study, probably because the $\delta^{15}\text{N}$ values of our samples were in a quite narrow range (1.65–10.85‰), or because other factors have currently a more pro-

nounced effect on the analytical quality. Furthermore, it should be pointed out that none of the participating labs used a single-point anchoring versus the working gas that Paul et al. (2007) showed to be most problematic.

3.4.2. Blanks

Because ambient air contains 78% dinitrogen, even small leaks in the analytical system may lead to significant blanks, and are thus of special concern for the accuracy of the $\delta^{15}\text{N}$ determination. While leaks at connections normally contribute to the background, leaks in the auto sampler may be a prominent source of blanks. A further source of blanks is dinitrogen adsorbed on the surface of the tin capsules (Hansen and Sommer, 2007). The contribution of blanks to the measured isotope ratio can be assessed by means of a simple mass balance:

$$\delta^{15}\text{N}_{\text{total}} * \text{area}_{\text{tot}} = \delta^{15}\text{N}_{\text{sample}} * \text{area}_{\text{sample}} + \delta^{15}\text{N}_{\text{blank}} * \text{area}_{\text{blank}} \quad (1)$$

where the $\delta^{15}\text{N}_{\text{blank}}$ -value is often assumed to be 0.0‰.

When assuming a constant blank area and a constant $\delta^{15}\text{N}_{\text{blank}}$ of 0.0‰, unnoticed blanks would lead to relatively depleted $\delta^{15}\text{N}$ values for samples that are ^{15}N enriched relative to atmospheric dinitrogen. According to the mass balance Eq. (1) above, the deviation from the true value should increase with increasing $\delta^{15}\text{N}$ values (assuming that $\delta^{15}\text{N}$ data are positive values as was the case in this study) and decreasing nitrogen content of the samples. The means of laboratory 2 are on average significantly lower than the grand average, but we observe no correlation of the bias with either the grand average or the peak height and thus exclude blanks as a source for these errors.

Only two laboratories (1 and 8) performed a blank correction during this exercise. The means of both laboratories are on average lower by 0.06‰ and 0.07‰ than grand averages and do not significantly differ from the means of the remaining laboratories (excluding laboratory 2). In order to exclude blanks from the auto-sampler, laboratory 5 has encapsulated the entire autosampler in an argon atmosphere to avoid blank problems. Laboratory 5 means are on average increased by 0.11‰, but still well within the range of the results reported by the other laboratories. In conclusion, our data provide no indication for a significant effect of potential blanks or blank correction procedures on the quality of the analytical results.

3.4.3. Influence of carbonate content/isobaric interferences

Werner and Brand (2001) reported that the CO_2 peak eluting from the GC column can tail into the nitrogen peak of following samples, and we have previously observed similar issues with some sample matrices which prove difficult to combust (e.g., glass fiber filters) and for which small peaks of CO_2 may elute after the main CO_2 peak. CO from the unimolecular decay of CO_2 provides an m/z 29/28 ratio equivalent to a $\delta^{15}\text{N}$ value of approximately +600‰, so that even traces of CO can significantly interfere with isotope analysis of N_2 . To overcome this problem, Werner and Brand (2001) recommended the removal of CO_2 from the sample gas stream before introduction into the IRMS. In our round robin test, no significant isobaric interferences from CO_2 were observed: out of the eight participating laboratories, three laboratories (4, 5 and 6) did not remove CO_2 from the sample gas stream. On average, the means from these three laboratories showed an offset of +0.09‰ and were well within the reported averages. The TC/TN ratio of the samples ranges from 7.7–348, and the bias of these three laboratories was not correlated with the total carbon content or the TC/TN ratios of the samples. We therefore exclude a significant effect of isobaric interferences from CO derived from sample CO_2 for results of this round robin test.

3.4.4. Nitrogen content and peak shape parameter

A box-and-whisker plot (Fig. 1) reveals a significantly higher variance for sample F compared to the others. This was confirmed by a Cochran's test for both corrected and uncorrected data and is related to the low nitrogen content (0.02%) of this sample. The standard deviations (SD) of samples calculated from uncorrected data related to total nitrogen content (N%) by a power law function (Fig. 2). Moreover, we observed an inverse correlation between the WLSD and the ratio of peak height (PH) to peak width (PW) (Fig. 3a), which was taken as a simple criterion for the peak shape. This correlation is more significant than those in which WLSD is related to PH or PW alone. The PW (and thus also the ratio PH/PW) mainly depends on the chromatographic conditions and carrier gas flow rate, as well as on the sample combustion.

In general, PW tends to increase with decreasing nitrogen content, because the combustion temperature will decrease somewhat with increasing sample weight, which may subsequently lead to a peak broadening. Low nitrogen contents in samples can often only partially be balanced by increasing the sample weight, and thus the PH will normally decrease with decreasing nitrogen content and at the same time the PW tends to increase.

Bräuer and Hahne (2005) have shown that ammonia adsorbed to silicates is not completely liberated with conventional EA-IRMS routines. This ammonia leads to analytical errors, because it is often enriched in ^{15}N relative to the organic fraction. The authors further found that the addition of V_2O_5 , which acts as an oxidation catalyst and also increases the temperature during flash combustion, completes the extraction of ammonia. In another study, Boyd (1997) showed that prolonged heating up to 1300°C removes ammonium quantitatively from different rocks. With respect to the present study, these results indicate that the liberation of bound ammonia and thus the determination of the N isotopic composition may be affected by subtle differences of the combustion conditions, which may explain the observed relation between the sample standard deviation and the peak shape. Furthermore, this

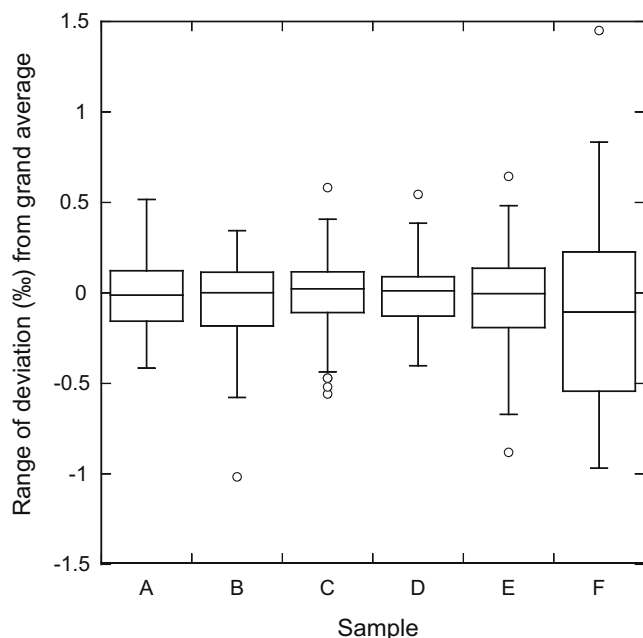


Fig. 1. Box-and-whisker plot showing results for each sample. Each box encloses 50% of the data with the median value of the variable displayed as a line. The top and bottom of the box mark the limits of $\pm 25\%$ of the variable population. The lines extending from the top and bottom of each box mark the minimum and maximum values within the data set that fall within an acceptable range. Any value outside of this range (outliers) is displayed as an individual point.

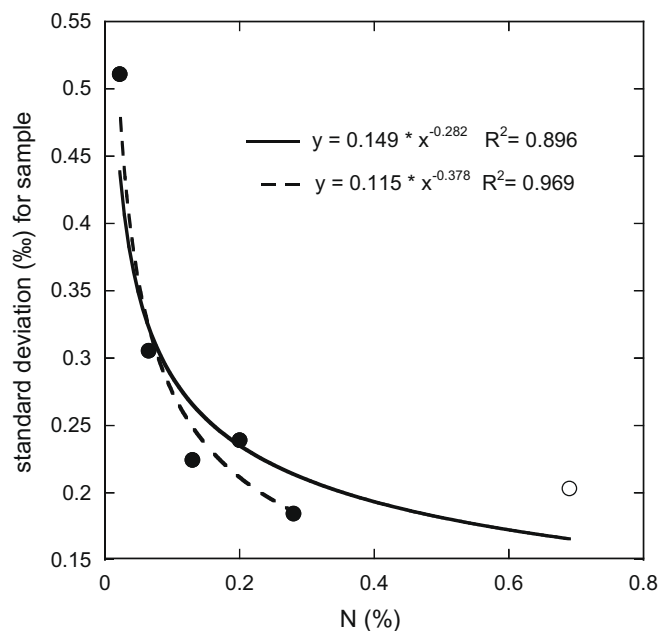


Fig. 2. Standard deviations of $\delta^{15}\text{N}$ determined for samples plotted against the N-concentration of the samples. Sample D (open symbol) has the highest N content, and has been excluded in calculating equation 2 (dashed line). Solid line represents fits to all data (including sample D).

effect may be more pronounced for samples with low nitrogen contents due to a larger fraction of lattice bound ammonia.

Beside these influences of materials, the peak shape itself may affect the peak integration. The following factors are directly related to the peak shape and may affect the WLSD:

- **Baseline noise:** With decreasing peak width, the contribution of the baseline noise to the measurement precision will be reduced.
- **Variations in integration limits:** The limits for peak integration are defined by the start and the end slope of the peak. Small variations of peak shape may therefore affect the fraction of the peak that will be integrated for calculation of the isotope ratios and thus the standard deviation of the $\delta^{15}\text{N}$ determination. With decreasing PH/PW ratios, the setting of the integration limits becomes more sensitive to subtle variations in peak shape and thus the standard deviation may increase.
- **Background correction:** The part of the peak before the start mark is attributed to the background, which is subtracted from the peak before calculation of the isotope ratios and thus affects the background correction. With increasing peak width, the background correction may introduce subtle errors.

Although narrow peaks are not a pre-requisite for a small WLSD, our data clearly show that improving the ratio PH/PW may substantially improve the WLSD.

The bias (deviation from the grand average) can be taken as a measure for systematic differences among the participating laboratories. Our data indicate that the bias is more likely related to the PW alone than to the ratio PH/PW. On the level of individual laboratories, a clear positive correlation between the bias and PH/PW is only observed for laboratories 2 and 3, while the data from all other laboratories behave indifferently; however, a significant correlation with PW is observed for laboratories 2, 3, 6 and 7 (Fig. 3b).

On the level of individual samples, we observe a positive trend of the bias with PH/PW, but a significant correlation is only observed for samples B and C after excluding the results of laboratory

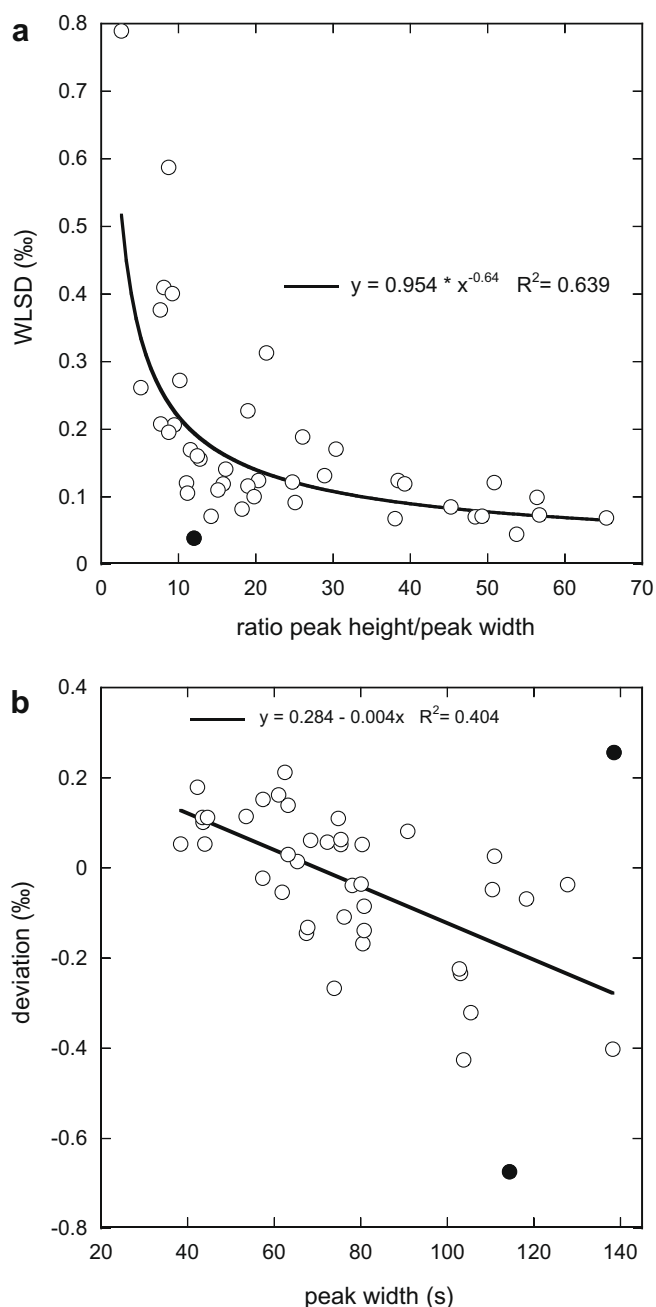


Fig. 3. Peak shape parameters peak height/peak width (A) and peak width (B) and their relation to within-laboratory standard deviation (A) and to measurement bias (B).

2, and for sample A after removing the results of laboratories 2 and 6. For samples D, E and F we observed no clear correlation. A correlation with PW alone is observed for sample B ($R^2 = 0.51$), sample C ($R^2 = 0.39$) and sample E ($R^2 = 0.69$).

The pooled dataset shows no significant correlation between the bias and the PH/PW ratio:

$$\text{BIAS (deviation)} = 0.002(\text{PH/PW}) - 0.088, n = 46, R^2 = 0.04$$

but a significant correlation of bias with PW (Fig. 3b):

$$\text{BIAS (deviation)} = -0.004 \text{ PW} + 0.284, n = 46, R^2 = 0.40.$$

We speculate that these differences are related to systematic variations of combustion conditions, of the integration limits and of the background correction as described above. In this case, the

bias should be related to systematic differences of the sample peak shape relative to the peak shape of the reference materials.

It is common practice to adjust the sample weight to obtain similar peak heights for all samples in a batch, and/or to match PH of samples to those of the reference gas. Therefore, relative variations in peak shape are reduced to variations in PW leading to the correlation observed here. This correlation is relatively weak, but it should be pointed out that the bias refers to the grand average of individual samples and not to their true isotopic composition. For this reason, the correlation cannot account for the observed differences in PW of individual samples. As a consequence, one may expect a significantly closer correlation for the true bias. Further, when accepting this correlation, it directly follows that the grand averages are associated with unknown systematic errors that are somehow related to the differences in the averaged PW of the samples. We thus propose that minimizing the differences in peak width between calibration standards and samples may reduce the bias.

3.4.5. Evaluation of the participating laboratories

In order to assess individual laboratory performance, we have pooled the deviations from the grand average of all measurements for samples A–E for each laboratory. Sample F was discarded from the evaluation of the participating laboratories because of its significantly elevated standard deviation, and because not all laboratories provided a full dataset for this sample. The bias (mean deviation from the grand average) and the corresponding SD are summarized in Table 4. Individual laboratory means for each sample are shown in Table 3.

A box-and-whisker plot (Fig. 4) shows the distribution of the pooled deviations for each laboratory. For laboratories 1, 2, 6, 7 and 8, the plot indicates a normal distribution of the measurement results, whereas for laboratory 4 and 5 (and to a lesser extent for laboratory 3), a skewed distribution is observed.

The bias is a measure for the overall trueness of the measurements (systematic errors) and the corresponding standard deviation is used as a measure for the within-laboratory precision (WLSD) respectively. The average bias ranged from -0.27‰ for laboratory 2 to $+0.11\text{‰}$ for laboratory 5. A Grubb's test denoted the deviation of laboratory 2 as a probable type 2 outlier (extreme values in the laboratory mean) in the laboratory trueness.

The WLSD of laboratories 4 and 8 were 0.41 and 0.32‰ , respectively, and were identified as outliers in the laboratory precision by means of a Cochran's test ($p = 0.05$). The WLSD of the remaining six laboratories averaged 0.15‰ and ranged from 0.10‰ (laboratory 5) to 0.20‰ (laboratory 6).

3.5. Expanded uncertainty estimates and recommendations

To the best of our knowledge, this is the first published round robin exercise on $\delta^{15}\text{N}$ in natural sediments that allows a comprehensive assessment of the current measurement uncertainty. Overall, we obtained a combined standard deviation (1 s) of 0.20‰ (0.28‰ when including sample F with an extremely low nitrogen content of 0.02%). If outliers are removed according to accepted procedures, the standard deviation would improve to 0.18‰ . However, one should keep in mind that most of these outliers would not have been recognized during routine operation. A more conservative estimate of overall measurement uncertainty based on 95% confidence interval would be 0.48‰ . This seems to be a reliable estimate of the current analytical precision of $\delta^{15}\text{N}$ determinations in sediments for samples with a nitrogen content $>0.07\%$. From the observed relationship between the PW and the bias (Fig. 3B), we conclude that the grand averages of the samples are associated with an unknown systematic error that is related to the differences in the averaged peak width of the samples in relation to the PW of

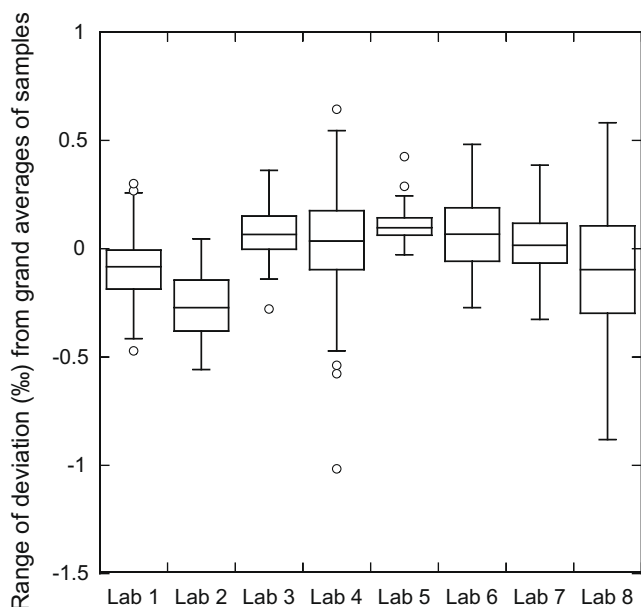


Fig. 4. Box-and-whisker plot of bias for each laboratory.

the reference materials. Currently it is not possible to assess the deviation of the grand average from the true value of the samples but there is some indication for an underestimation of the true values with decreasing nitrogen content, which is in line with an incomplete liberation of ammonia fixed to siliceous material.

This expanded measurement uncertainty of 0.48‰ for marine sediments is much better than that recently reported for a single cod tissue sample analyzed by eight laboratories (Mill et al., 2008). From the variances reported by these authors, we have deduced an expanded measurement uncertainty of 1.59‰ for their study. Similarly, the difference among laboratories (bias) was much better in our study (laboratory average from $-0.27‰$ to $+0.11‰$; averaged bias after exclusion of sample F. The maximum ranges of mean biases were 0.81‰ for sample B and 0.93‰ for sample F, lower than the range of 2.7‰ in Mill et al. (2008). The reasons for these differences are unknown and cannot be assessed, because Mill et al. (2008) did not report the analytical procedures, instruments and other essential details. Although we agree that between-laboratory differences should be acknowledged, we believe that current laboratory procedures are sufficient to attain high accuracies (i.e. limited bias) and WLSD of about 0.15‰. Moreover, our results provide some promising clues for further improvement of the quality of $\delta^{15}\text{N}$ determinations. Although variations in sediment $\delta^{15}\text{N}$ can be substantial in some cases (e.g., Gaye-Haake et al., 2005; Voß et al., 2005), even millenium scale variations in ocean sedimentary $\delta^{15}\text{N}$ records can fluctuate within a narrow (1–2‰) range (Altabet, 2007), rendering very precise and accurate determinations crucial to correctly document and interpret nitrogen stable isotope signatures in the sedimentary record.

Different authors have recommended a linearity correction via the PH to account for differences of the partial pressure of nitrogen gas in the ion source. The linearity of modern IRMS-instruments is normally better than 0.04‰/V. Our results indicate a more prominent effect of the peak shape on the quality of the $\delta^{15}\text{N}$ determination than the peak height alone. We recommend minimizing the differences in PW between calibration standards and samples in order to reduce potential bias. To our experience, these differences are often related to the sample weight and can to some extent be outbalanced by optimizing the combustion conditions. With increasing PW the contribution of the baseline noise to the total

signal will increase and thus increase the standard deviation. This contribution is counterbalanced by the PH, leading to the observed correlation between the ratio of PH/PW and the WLSD. Thus, minimizing the PW appears to be a crucial step for minimizing the WLSD. Although narrow peaks are not a pre-requisite for a good WLSD, our data clearly show that improving the ratio PH/PW may substantially improve the WLSD. Systematic use of certified reference materials with matrices close to those of the samples and participation in robin ring tests will eventually result in more accurate and more precise dataset for use in various environmental disciplines.

4. Conclusions

- Data for samples with nitrogen content above 0.07% were surprisingly good with an expanded measurement uncertainty of 0.48‰; the difference among laboratories (bias) in our study ranged from $-0.27‰$ to $+0.11‰$ and was much better than the bias reported previously.
- The expanded uncertainty may be improved to 0.36‰ after improving individual quality of results (laboratory 2, 4 and 8).
- Improving peak shape may be a way to improve the measurement precision of individual laboratories.
- It is not possible to assess the deviation of the grand average from the true value of the samples, but there is some indication for an underestimation of the true values with decreasing nitrogen content.
- We see an urgent need for certified natural reference materials that are in the $\delta^{15}\text{N}$ range of naturally occurring marine sediments.

Acknowledgements

The study was initiated as part of EUROCORES project MERF (04 ECLIM FP35) and funded under contract DFG Em-37/20. Support is acknowledged from the Fund for Scientific Research (FWO-Vlaanderen, contracts G.0632.06 and G.0395.07).

References

- Altabet, M.A., 2007. Constraints on oceanic N balance/imbalance from sedimentary N-15 records. *Biogeosciences* 4 (1), 75–86.
- Boyd, S.R., 1997. Determination of the ammonium content of potassic rocks and minerals by capacitance manometry: a prelude to the calibration of FTIR microscopes. *Chemical Geology* 137, 57–66.
- Bräuer, K., Hahne, K., 2005. Methodical aspects of the ^{15}N -analysis of Precambrian and Paleozoic sediments rich in organic matter. *Chemical Geology* 218, 361–368.
- DeNiro, M.J., Epstein, S., 1981. Influence of diet on the distribution of nitrogen isotopes in animals. *Geochimica et Cosmochimica Acta* 45, 341–351.
- Fry, B., 2007. *Stable Isotope Ecology*. Springer, Berlin, 317 pp.
- Gaye-Haake, B., Lahajnar, N., Emeis, K.-C., Rixen, T., Unger, D., Schulz, H., Ramaswamy, V., Paropkari, A.L., Gupta, M.V.S., Ittekkot, V., 2005. Stable nitrogen isotopic ratios of sinking particles and sediments from the northern Indian Ocean. *Marine Chemistry* 96, 243–255.
- Hansen, T., Sommer, U., 2007. Increasing the sensitivity of $\delta^{13}\text{C}$ and $\delta^{15}\text{N}$ abundance measurements by a high sensitivity elemental analyzer connected to an isotope ratio mass spectrometer. *Rapid Communications in Mass Spectrometry* 21 (3), 314–318.
- Jardine, T., Cunjak, R., 2005. Analytical error in stable isotope ecology. *Oecologia* 144, 528–533.
- Jaschinski, S., Brepohl, D.C., Sommer, U., 2008. Carbon sources and trophic structure in an eelgrass *Zostera marina* bed, based on stable isotope and fatty acid analyses. *Marine Ecology Progress Series* 358, 103–114.
- Kennedy, N., Connolly, J., Clipson, N.J.W., 2005. Impact of lime, nitrogen and plant species on fungal community structure in grassland microcosms. *Environmental Microbiology* 7, 780–788.
- McClelland, J.W., Valiela, I., 1998. Linking nitrogen in estuarine producers to land-derived sources. *Limnology Oceanography* 43, 577–585.
- Middelburg, J.J., Herman, P.M.J., 2007. Organic matter processing in tidal estuaries. *Marine Chemistry* 106, 127–147.

- Mill, A.C., Sweeting, C.J., Barnes, C., Al-Habsi, S.H., MacNeil, A., 2008. Mass-spectrometer bias in stable isotope ecology. *Limnology Oceanography Methods* 6, 34–39.
- Owens, N.J.P., 1987. Natural variations in ^{15}N in the marine environment. *Advances in Marine Biology* 24, 389–451.
- Paul, D., Skrzypek, G., Forizs, I., 2007. Normalization of measured stable isotopic compositions to isotope reference scales – a review. *Rapid Communications in Mass Spectrometry* 21, 3006–3014.
- Peters, K.E., Sweeney, R.E., Kaplan, I.R., 1978. Correlation of carbon and nitrogen stable isotope ratios in sedimentary organic matter. *Limnology Oceanography* 23, 598–604.
- Preston, T., Owens, N.J.P., 1983. Interfacing and automatic an elemental analyser with an isotope ratio mass spectrometer: the potential for fully automated total nitrogen and nitrogen-15 analysis. *Analyst* 108, 971–977.
- Sweeney, R.E., Kaplan, I.R., 1980. Natural abundances of ^{15}N as a source indicator for near-shore marine sedimentary and dissolved nitrogen. *Marine Chemistry* 9, 81–90.
- Verkouteren, R.M., 1999. Preparation, characterization, and value assignment of carbon dioxide isotopic reference materials: RMs 8562, 8563, and 8564. *Analytical Chemistry* 71 (20), 4740–4746.
- Voß, M., Emeis, K.-C., Hille, S., Neumann, T., Dippner, J., 2005. The nitrogen cycle of the Baltic Sea from an isotopic perspective. *Global Biogeochemical Cycles* 19 (3). doi:10.1029/2004GB002338.
- Wada, E., Terazaki, M., Kabaya, Y., Nemoto, T., 1987. ^{15}N and ^{13}C abundances in the Antarctic Ocean with emphasis on the biogeochemical structure of the food web. *Deep-Sea Research* 34, 829–841.
- Werner, R.A., Brand, W.A., 2001. Referencing strategies in stable isotope ratio analysis. *Rapid Communications in Mass Spectrometry* 15, 501.